

STIC-Biotech/ChemLib

78418

Paper 20 Attach.

From: Loeb, Bronwen
Sent: Tuesday, October 22, 2002 3:15 PM
To: STIC-Biotech/ChemLib
Subject: sequence search request 09/717,321

09/717,321
Rothberg et al
filed Nov. 20, 2000

Please search SEQ ID NO. 1 in all databases.

Thank you.

Bronwen Loeb, PhD

AU 1636
703-605-1197
CM1 11D-16
Mailbox 11E-12

POINT OF CONTACT:
PAUL SCHULWITZ
TECHNICAL INFO. SPECIALIST
CM1 6B06 TEL. (703) 305-1954

RECEIVED
OCT 22 2002
STIC

Start 10/24 10
Fin 10/28 10

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 26, 2002, 03:02:20 ; Search time 1600 Seconds
(without alignments)
1037.578 Million cell updates/sec

Title: US-09-717-321a-1
Perfect score: 123
Sequence: 1 caattgaaagagtttttc.....ctcacagaccacaaagtacc 123

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	122	99.2	277	10 BF420446	BF420446 UI-R-BJ2-
2	122	99.2	292	9 AI598992	AI598992 EST250695
3	122	99.2	316	9 AI412434	AI412434 EST240733
4	122	99.2	404	10 BF549054	BF549054 UI-R-A0-a
5	122	99.2	405	9 BE104696	BE104696 UI-R-BX0-
6	122	99.2	407	9 AI009011	AI009011 EST203462
7	122	99.2	429	9 AI059212	AI059212 UI-R-C1-1
8	122	99.2	455	9 AI012196	AI012196 EST206647
9	122	99.2	463	10 BF548903	BF548903 UI-R-A0-a
10	122	99.2	465	9 AA899757	AA899757 UI-R-E0-d
11	122	99.2	477	9 AA520555	AA520555 UI-R-BJ0p
12	122	99.2	489	9 AA859879	AA859879 UI-R-E0-c
13	122	99.2	493	10 BF554834	BF554834 UI-R-E0-c
14	122	99.2	522	10 C06842	C06842 C06842 Rat
15	122	99.2	565	10 BI273986	BI273986 UI-R-CW0-
16	122	99.2	615	10 BG380448	BG380448 UI-R-CT0-
17	122	99.2	639	9 BE111691	BE111691 UI-R-BJ1-

18	122	99.2	687	10 BM389059	BM389059 UI-R-DZ0-
19	120.4	97.9	316	9 AI409795	AI409795 EST238088
20	120.4	97.9	557	10 BG665005	BG665005 DRABXA06
21	118.8	96.6	517	9 AA859940	AA859940 UI-R-E0-C
22	118.8	96.6	553	9 AA799542	AA799542 EST189039
23	118.8	96.6	638	10 BI283790	BI283790 UI-R-CW0s
24	117.2	95.3	294	9 AI236740	AI236740 EST233302
25	117.2	95.3	469	10 BI276611	BI276611 UI-R-CX0-
26	115.6	94.0	431	9 AI410870	AI410870 EST239163
27	115.6	94.0	446	10 BI286743	BI286743 UI-R-CT0s
28	112.4	91.4	256	9 AI234716	AI234716 EST231278
29	110.8	90.1	427	9 AI178082	AI178082 EST221741
30	107.6	87.5	225	9 BE136269	BE136269 ug28f02.x
31	107.6	87.5	240	9 AW323013	AW323013 uc56c06.x
32	107.6	87.5	283	9 AW122670	AW122670 UI-M-BH2.
33	107.6	87.5	288	9 AV128839	AV128839 AV128839
34	107.6	87.5	294	10 W13019	W13019 mb22c12.r1
35	107.6	87.5	298	9 AA940186	AA940186 ua29f01.r
36	107.6	87.5	299	9 BB749463	BB749463 BB749463
37	107.6	87.5	309	9 AA265870	AA265870 mz70h02.r
38	107.6	87.5	317	9 AA396854	AA396854 mr38b07.r
39	107.6	87.5	340	9 AA388802	AA388802 vb25c09.r
40	107.6	87.5	356	9 AA163447	AA163447 mr27g12.r
41	107.6	87.5	363	9 BB686346	BB686346 BB686346
42	107.6	87.5	376	10 BM200395	BM200395 C0210P10-
43	107.6	87.5	378	9 AA461828	AA461828 vf55c06.r
44	107.6	87.5	380	9 AA213220	AA213220 mw80e02.r
45	107.6	87.5	386	9 BB685587	BB685587 BB685587

ALIGNMENTS

RESULT 1

BF420446

LOCUS

DEFINITION

BF420446

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

277 bp mRNA linear EST 28-NOV-2000
UI-R-BJ2-bpw-e-11-0-UI.s1 UI-R-BJ2 Rattus norvegicus CDNA clone
UI-R-BJ2-bpw-e-11-0-UI 3', mRNA sequence.

BF420446.1 GI:11408475

EST.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 277)

Bonaldi, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to identify it as a clone from the

normalized atrium at 16.5 dpc library cDNA Library Preparation:

M.B. Soares Lab Clone distribution: clones will be available

through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA-Yes.

Location/Qualifiers

1. 277

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

FEATURES

Source

/db_xref="taxon:10116"
 /clone="UI-R-BJ2-bpw-e-11-0-UI"
 /clone_lib="UI-R-BJ2"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ2
 library is a subtracted library derived from the following
 tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc,
 atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15
 dpc, AV canal at 15 dpc. For a detailed description of
 the library from which this clone was derived, please
 visit our web site at ratest.eng.uiowa.edu. The
 subtraction has been previously described in (Bonaldi,
 Lennon and Soares, Genome Research 6:791-806, 1996)
 TAG_LIB=UI-R-BJ2
 TAG_ISSUE=atrium at 16.5 dpc
 TAG_SEQ=GATTC
 TAG_SEQ=ATTC

BASE COUNT 85 a 58 c 69 g 65 t
 ORIGIN

Query Match 99.2%; Score 122; DB 10; Length 277;
 Best Local Similarity 100.0%; Pred. No. 3e-30;
 Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CAATTGAAAAAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTCTTCCAGTGAG 60
 DB 66 CAATTGAAAAAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTCTTCCAGTGAG 125
 QY 61 TTAGTTGTACAGACGGCGTTAGCCTAGCGCTTGACAGACCTTCACAGACCCAAAGGT 120
 DB 126 TTAGTTGTACAGACGGCGTTAGCCTAGCGCTTGACAGACCTTCACAGACCCAAAGGT 185
 QY 121 AC 122
 DB 186 AC 187

RESULT 2
 LOCUS A1598992 292 bp mRNA linear EST 21-APR-1999
 DEFINITION E2750695 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
 REMEG05 3' end, mRNA sequence.
 ACCESSION A1598992
 VERSION A1598992.1 GI:4608040
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 292)
 Lee, N.H., Glodex, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Kerlavage, A.R. and Adams, M.D.
 Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
 Gene Index
 Unpublished (1998)
 Contact: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhleest@igf.org
 Seq primer: M13-21.

FEATURES
 Location/Qualifiers
 1..292
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /clone="REM05"
 /clone_lib="Normalized rat embryo, Bento Soares"
 /dev_stage="embryo 8, 12, 18 dpc"
 /note="vector: pRT3Pac; Site_1: EcoRI; Site_2: NotI"
 89 a 60 c 70 g 73 t

BASE COUNT
 ORIGIN

Query Match 99.2%; Score 122; DB 9; Length 292;
 Best Local Similarity 100.0%; Pred. No. 3e-30;
 Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CAATTGAAAAAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTCTTCCAGTGAG 60
 DB 97 CAATTGAAAAAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTCTTCCAGTGAG 156
 QY 61 TTAGTTGTACAGACGGCGTTAGCCTAGCGCTTGACAGACCTTCACAGACCCAAAGGT 120
 DB 157 TTAGTTGTACAGACGGCGTTAGCCTAGCGCTTGACAGACCTTCACAGACCCAAAGGT 216
 QY 121 AC 122
 DB 217 AC 218

RESULT 3
 LOCUS A1412434 316 bp mRNA linear EST 09-FEB-1999
 DEFINITION E1240733 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
 RBRDO65 3' end, mRNA sequence.
 ACCESSION A1412434
 VERSION A1412434.1 GI:4255938
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 316)
 Lee, N.H., Glodex, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Kerlavage, A.R. and Adams, M.D.
 Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
 Gene Index
 Unpublished (1998)
 Contact: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhleest@igf.org
 Seq primer: M13-21.

FEATURES
 Location/Qualifiers
 1..316
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /clone="RBRDO65"
 /clone_lib="Normalized rat brain, Bento Soares"
 /note="Organ: brain; Vector: pRT3Pac; Site_1: EcoRI;
 Site_2: NotI"
 98 a 62 c 76 g 80 t

BASE COUNT 98 a 62 c 76 g 80 t
 ORIGIN

Query Match 99.2%; Score 122; DB 9; Length 316;
 Best Local Similarity 100.0%; Pred. No. 3.1e-30;
 Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CAATTGAAAAAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTCTTCCAGTGAG 60
 DB 97 CAATTGAAAAAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTCTTCCAGTGAG 156
 QY 61 TTAGTTGTACAGACGGCGTTAGCCTAGCGCTTGACAGACCTTCACAGACCCAAAGGT 120
 DB 157 TTAGTTGTACAGACGGCGTTAGCCTAGCGCTTGACAGACCTTCACAGACCCAAAGGT 216
 QY 121 AC 122
 DB 217 AC 218

RESULT 4

BF549054/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BF549054 404 bp mRNA linear EST 11-DEC-2000
UI-R-A0-ag-b-01-0-UI.r1 UI-R-A0 Rattus norvegicus cDNA clone
UI-R-A0-ag-b-01-0-UI 5', mRNA sequence.
BF549054
BF549054.1 GI:11640209
EST.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

1 (bases 1 to 404)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msobres@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
The following repetitive elements were found in this cDNA sequence:
1-22, >AT-richflow_complexity
Seq primer: M13 Forward
POLYA=yes.

FEATURES
source
Location/Qualifiers
1..404
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-A0-ag-b-01-0-UI"
/clone_lib="UI-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; This library
consists of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung, brain
liver, kidney, heart, spleen, ovary, and muscle. The tag
is a string of 3-5 nucleotides present between the Not I
site and the oligo-dT track which allows identification of
the library of origin of a clone within the mixture."
BASE COUNT 92 a 98 c 83 g 131 t
ORIGIN
Query Match 99.2%; Score 122; DB 10; Length 404;
Best Local Similarity 100.0%; Pred. No. 3.3e-30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAAAGTTTCTCTAGTGGTCGAAAGCCCAACACTGTGTCTGCCAGTGAG 60
|||||
DB 368 CAATGAAAAAGTTTCTCTAGTGGTCGAAAGCCCAACACTGTGTCTGCCAGTGAG 309
|||||
QY 61 TTAGTGTGTACAGAACGCGGTAGCCTAGCGCTTGACAGAACCTCAGACACCCAAAGGT 120
|||||
DB 308 TTAGTGTGTACAGAACGCGGTAGCCTAGCGCTTGACAGAACCTCAGACACCCAAAGGT 249
|||||
QY 121 AC 122
||
DB 248 AC 247

RESULT 5
BE104696
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE104696 405 bp mRNA linear EST 13-JUN-2000
UI-R-BX0-ars-e-01-0-UI.s1 UI-R-BX0 Rattus norvegicus cDNA clone
UI-R-BX0-ars-e-01-0-UI 3', mRNA sequence.
BE104696
BE104696.1 GI:8496796
EST.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

1 (bases 1 to 404)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msobres@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
The following repetitive elements were found in this cDNA sequence:
1-22, >AT-richflow_complexity
Seq primer: M13 Forward
POLYA=yes.

FEATURES
source
Location/Qualifiers
1..405
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BX0-ars-e-01-0-UI"
/clone_lib="UI-R-BX0"
/dev_stage="embryonic 13 dpc"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; The UI-R-BX0
library is derived from 13 dpc whole embryo tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
rategen.uiowa.edu.
TAG-SQ=None found"
BASE COUNT 131 a 76 c 87 g 110 t
ORIGIN
Query Match 99.2%; Score 122; DB 9; Length 405;
Best Local Similarity 100.0%; Pred. No. 3.3e-30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAAAGTTTCTCTAGTGGTCGAAAGCCCAACACTGTGTCTGCCAGTGAG 60
|||||
DB 117 CAATTGAAAAAGTTTCTCTAGTGGTCGAAAGCCCAACACTGTGTCTGCCAGTGAG 176
|||||
QY 61 TTAGTGTGTACAGAACGCGGTAGCCTAGCGCTTGACAGAACCTCAGACACCCAAAGGT 120
|||||
DB 177 TTAGTGTGTACAGAACGCGGTAGCCTAGCGCTTGACAGAACCTCAGACACCCAAAGGT 236
|||||
QY 121 AC 122
||
DB 237 AC 238

RESULT 6
AI009011
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AI009011 407 bp mRNA linear EST 25-JAN-1999
EST203462 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
RMBH20 3' end, mRNA sequence.
AI009011
AI009011.1 GI:3222843
EST.
Rattus sp.
Rattus sp.

REFERENCE	1 (bases 1 to 429)
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
COMMENT	Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: msoares@blue.weeg.uiowa.edu The sequence tag present in the cDNA between the NotI site and the oligo-dr track served to identify it as a clone from the normalized

RESULT 8.	AI012196	AI012196	AI012196	455 bp	mRNA	linear	EST 15-JUN-1998
LOCUS	EST206647	Normalized rat placenta,	Bento Soares	Rattus sp.	CDNA		
DEFINITION	clone RPLAT85	3' end,	mRNA	sequence.			
ACCESSION	AI012196						
VERSION	AI012196.1	GI:3226028					
KEYWORDS	EST.						
SOURCE	Rattus sp.						
ORGANISM	Rattus sp.						
	Eukaryota;	Metazoa;	Chordata;	Craniata;	Vertebrata;	Euteleostomi;	
	Mammalia;	Eutheria;	Rodentia;	Sciurognathi;	Muridae;	Murinae;	
	Rattus.						
REFERENCE	1	(bases 1 to 455)					

AUTHORS Lee, N.H., Glodok, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R., and Adams, M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES Location/Qualifiers
source 1. 455
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone_lib="RPLA185"
/note="Normalized rat placenta, Bento Soares"
Site_2: NotI
156 a 92 c 102 g 105 t

BASE COUNT 156 a 92 c 102 g 105 t

ORIGIN
Query Match 99.2%; Score 122; DB 9; Length 455;
Best Local Similarity 100.0%; Pred. No. 3.4e-30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAGTTTCTTCTAGTGGTGGAAAGGCCCAACACTGTGTCTTCCAGTGAG 60
Db 91 CAATTGAAAAGTTTCTTCTAGTGGTGGAAAGGCCCAACACTGTGTCTTCCAGTGAG 150
QY 61 TTAGGTGTACAGACGGCGTTAGCTAGCGCTTACAGAACTCAGACACCAAGGT 120
Db 151 TTAGGTGTACAGACGGCGTTAGCTAGCGCTTACAGAACTCAGACACCAAGGT 210
QY 121 AC 122
Db 211 AC 212

RESULT 9
BF548903/c 463 bp mRNA linear EST 11-DEC-2000
LOCUS UI-R-A0-ag-b-08-0-UI.r1 UI-R-A0 Rattus norvegicus cDNA clone
DEFINITION UI-R-A0-ag-b-08-0-UI 5', mRNA sequence.
ACCESSION BF548903
VERSION BF548903.1 GI:11640010
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 463)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone Distribution:
This clone is also available through Research Genetics (www.resgen.com)
LNL (info@lml.lnl.gov). IMAGE ID= I791813
Seq primer: M13 Forward.
Location/Qualifiers
1. 463

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="UI-R-A0-ag-b-08-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; This library
consists of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung, brain
liver, kidney, heart, spleen, ovary, and muscle. The tag
is a string of 3-5 nucleotides present between the Not I
site and the oligo-dT track which allows identification of
the library of origin of a clone within the mixture."

BASE COUNT 109 a 106 c 93 g 155 t

ORIGIN
Query Match 99.2%; Score 122; DB 10; Length 463;
Best Local Similarity 100.0%; Pred. No. 3.4e-30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAGTTTCTTCTAGTGGTGGAAAGGCCCAACACTGTGTCTTCCAGTGAG 60
Db 368 CAATTGAAAAGTTTCTTCTAGTGGTGGAAAGGCCCAACACTGTGTCTTCCAGTGAG 309
QY 61 TTAGGTGTACAGACGGCGTTAGCTAGCGCTTACAGAACTCAGACACCAAGGT 120
Db 308 TTAGGTGTACAGACGGCGTTAGCTAGCGCTTACAGAACTCAGACACCAAGGT 249
QY 121 AC 122
Db 248 AC 247

RESULT 10
AA899757 465 bp mRNA linear EST 05-FEB-1999
LOCUS UI-R-E0-da-b-11-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
DEFINITION UI-R-E0-da-b-11-0-UI 3' similar to gi|464185|dbj|D25274|HUMPO25T9
Homo sapiens mRNA, clone:PO2ST9, mRNA sequence.
ACCESSION AA899757
VERSION AA899757.1 GI:4232251
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 465)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Apr 7, 1998 this sequence version replaced gi:3035111.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone
distribution: clones will be available through Research Genetics
Seq primer: M13 Forward.
Location/Qualifiers
1. 465

FEATURES source
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="UI-R-E0-da-b-11-0-UI"
/dev_stage="adult"

```

/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: NotI; Site 2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dr track which allows
identification of the library of origin of a clone within
the mixture."
BASE COUNT      155 a      99 g      118 t
ORIGIN

```

```

Query Match      99.2%; Score 122; DB 9; Length 465;
Best Local Similarity 100.0%; Pred. No. 3.4e-30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CAATTGAAAAAGTTTCTAGTGTGCGAAGGCCCAACACTGTGTCTTCCAGTGAG 60
    |||||||
Db 112 CAATTGAAAAAGTTTCTAGTGTGCGAAGGCCCAACACTGTGTCTTCCAGTGAG 171
    |||||||
61 TTAGTTGTACAGAACGGCGTTAGCAGTGTGACAGAACCTTCACAGACCCAAAGGT 120
    |||||||
Db 172 TTAGTTGTACAGAACGGCGTTAGCAGTGTGACAGAACCTTCACAGACCCAAAGGT 231
    |||||||

```

```

QY 121 AC 122
Db 232 AC 233

```

```

RESULT 11
AW520555
LOCUS
DEFINITION
UI-R-BJ0p-afx-g-09-0-UI-s1 UI-R-BJ0p Rattus norvegicus cDNA clone
UI-R-BJ0p-afx-g-09-0-UI 3', mRNA sequence.
ACCESSION
AW520555
VERSION
AW520555.1 GI:7162933
KEYWORDS
EST.
SOURCE
Norway rat.

```

```

ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 477)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL
MEDLINE
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dr track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dr track served to identify it as a clone from the
normalized AV canal at 15 dpc library cDNA library preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

```

```

FEATURES
SOURCE

```

```

Location/Qualifiers
1..477

```

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ0p-afx-g-09-0-UI"
/clone_lib="UI-R-BJ0p"

```

```

/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ0p
library is a subtracted library derived from the UI-R-AA1,
UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
UI-R-AG1 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
canal at 15 dpc, and ventricle at 13 dpc. The tag is a
string of 5-6 nucleotides present between the Not I site
and the oligo-dr track. The library was constructed as
described by Bonaldo, Lennon and Soares, Genome Research
6: 791-806, 1996.
TAG_LIB=UI-R-BJ0p
TAG_ISSUE=AV canal at 15 dpc
TAG_SEQ=GAAGG*
BASE COUNT      157 a      103 g      124 t
ORIGIN

```

```

Query Match      99.2%; Score 122; DB 9; Length 477;
Best Local Similarity 100.0%; Pred. No. 3.5e-30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CAATTGAAAAAGTTTCTAGTGTGCGAAGGCCCAACACTGTGTCTTCCAGTGAG 60
    |||||||
Db 112 CAATTGAAAAAGTTTCTAGTGTGCGAAGGCCCAACACTGTGTCTTCCAGTGAG 171
    |||||||
61 TTAGTTGTACAGAACGGCGTTAGCAGTGTGACAGAACCTTCACAGACCCAAAGGT 120
    |||||||
Db 172 TTAGTTGTACAGAACGGCGTTAGCAGTGTGACAGAACCTTCACAGACCCAAAGGT 231
    |||||||

```

```

QY 121 AC 122
Db 232 AC 233

```

```

RESULT 12
AA859879
LOCUS
DEFINITION
UI-R-E0-cc-c-06-0-UI-s1 UI-R-E0 Rattus norvegicus cDNA clone
UI-R-E0-cc-c-06-0-UI 3' similar to dbj|D25274|HUMPO2ST9 Human
randomly sequenced mRNA, mRNA sequence.
ACCESSION
AA859879
VERSION
AA859879.1 GI:4230416
KEYWORDS
EST.
SOURCE
Norway rat.

```

```

ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 489)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL
MEDLINE
COMMENT
On Mar 10, 1998 this sequence version replaced gi:2949399.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dr track served to identify it as a clone from the normalized
adult 18-Day-Embryo library. cDNA Library Preparation: M. Fatima
Bonaldo, Ph.D. Clone distribution: clones will be available through
Research Genetics This clone is also available through the
I.M.A.G.E. Consortium at LNL (info@image.llnl.gov). IMAGE
ID=1770064
Seq primer: M13 Forward

```

FEATURES source

Best Local Similarity 100.0%; Pred. No. 3.5e-30;
Matches 122; Conservative 0; Mismatches 0;

Best local similarity 100.00, Rec. NO. 3.5e 30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

	OY	1	CAATTGAAAAAGTTGTCTAGTGGTCGAAGGCCCAACACTGTGTTCTTGCCAGTAG	60
	Ddb	69	CARTTGAAAAAGTTGTCTAGTGGTCGAAGGCCCAACACTGTGTTCTTGCCAGTAG	128
	OY	61	TTAGTGTGTACAGAACGGCGTTAGCACTAGCGCTTGACAACTCACAAGCCCCAAGGT	120
	Ddb	129	TTAGTGTGTACAGAACGGCGTTAGCACTAGCGCTTGACAACTCACAAGCCCCAAGGT	188
	OY	121	AC	122
	Ddb	189	AC	190

[illegible]

REFERENCE	1 (bases 1 to 565)
AUTHORS	Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477

Contact: Soares, MB
 Program for Rat Gene
 Discovery and Mapping
 University of Iowa
 451 Eckstein Medical
 Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565

Tel: 519.335.8001
 Email: mscoares@blue.wesg.uwoa.edu
 The sequence contained an oligo-dr track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dr track served to identify it as a clone from the normalized rat aorta pool library cDNA library preparation: M.B. Soares Lab Clone distribution. Clones will be available through Research Genetics (www.resgen.com)

```

FEATURES
source
POLYMERES.
Location/Qualifiers
1. .565
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="UI-R-CW0-bwb-c-12-0-UI"
/clone_lib="UI-R-CW0"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; The UI-R-CW0
library is a normalized library constructed from the
following rat aorta tissues: embryonic day 19, embryonic
day 21, adult day 1, adult day 12, adult day 75, adult day
200. For a detailed description of the library from which
this clone was derived, please visit our web site at
rattest.eng.ullowa.edu. The subtraction has been previously
described in (Bonaldio, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=UI-R-CW0
TAG_TISSUE=rat aorta pool

```

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	31.4	25.5	605	4	US-09-385-982-483	Sequence 483, Appl
2	25.6	20.8	3033	3	US-08-462-080B-1	Sequence 1, Appl
3	25.6	20.8	3033	3	US-08-462-090-1	Sequence 1, Appl
4	25.6	20.8	3033	3	US-08-463-461-1	Sequence 1, Appl
c 5	25.2	20.5	49272	1	US-08-614-770A-1	Sequence 1, Appl
6	25	20.3	550	3	US-08-705-771-1	Sequence 1, Appl
7	25	20.3	786	5	PCT-US95-08295-1	Sequence 1, Appl
8	24.6	20.0	3905	1	US-08-452-083-1	Sequence 1, Appl
9	24.6	20.0	80595	4	US-09-078-294-3	Sequence 3, Appl
c 10	23.4	19.0	228	1	US-08-250-958-3	Sequence 3, Appl
c 11	23.4	19.0	231	5	PCT-US95-00605-2	Sequence 2, Appl
12	23.4	19.0	588	4	US-09-385-982-133	Sequence 133, Appl
c 13	23.4	19.0	741	3	US-07-927-391-25	Sequence 25, Appl
c 14	23.2	18.9	576	1	US-08-086-428B-11	Sequence 11, Appl
c 15	23.2	18.9	576	2	US-08-468-570-11	Sequence 11, Appl
c 16	23.2	18.9	576	2	US-08-290-665A-11	Sequence 11, Appl
c 17	23.2	18.9	576	5	PCT-US95-10398-11	Sequence 11, Appl
c 18	23.2	18.9	3216	1	US-08-828-007-1	Sequence 1, Appl
c 19	23.2	18.9	3385	2	US-08-405-392-1	Sequence 1, Appl
c 20	23.2	18.9	3385	3	US-08-487-691-1	Sequence 1, Appl
c 21	23.2	18.9	3385	3	US-08-666-221B-3	Sequence 3, Appl
c 22	23.2	18.9	3385	3	US-08-666-221B-9	Sequence 9, Appl
c 23	23.2	18.9	4134	2	US-08-817-090B-1	Sequence 1, Appl
24	23.2	18.9	4137	2	US-08-817-090B-3	Sequence 3, Appl
c 25	23.2	18.9	4322	2	US-08-537-342-1	Sequence 1, Appl
26	23.2	18.9	5162	2	US-08-916-917-13	Sequence 13, Appl
27	23.2	18.9	5162	3	US-09-225-170-13	Sequence 13, Appl

```

1  GENERAL INFORMATION:
2  APPLICANT: Fowler, Timothy
3  APPLICANT: Barnett, Christopher C.
4  APPLICANT: Shoemaker, Sharon
5  TITLE OF INVENTION: Saccharification of Cellulose by Cloning
6  TITLE OF INVENTION: and Amplification of the Beta-glucosidase Gene of
7  TITLE OF INVENTION: Trichoderma Reesei
8  NUMBER OF SEQUENCES: 4
9  CORRESPONDENCE ADDRESS:
10 ADDRESS: Burns, Doane, Swecker & Mathis
11 CITY: George Mason Building, 699 Prince St.
12 STATE: Alexandria
13 COUNTRY: Virginia
14 ZIP: 22313-1404
15
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: Patentin Release #1.0, Version #1.25
21
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/462,090
24 FILING DATE: 05-JUN-1995
25 CLASSIFICATION: 435
26
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: US 07/625,140
29 FILING DATE: 10-DEC-1990
30 ATTORNEY/AGENT INFORMATION:
31 NAME: Dillahunty, T. Gene
32 REGISTRATION NUMBER: 25,423
33 REFERENCE/DOCKET NUMBER: 010055-056
34
35 TELECOMMUNICATION INFORMATION:
36 TELEPHONE: 415-854-7400
37 TELEFAX: 415-854-8275
38
39 INFORMATION FOR SEQ ID NO: 1:
40 SEQUENCE CHARACTERISTICS:
41 LENGTH: 3033 base pairs
42 TYPE: nucleic acid
43 STRANDEDNESS: single
44 TOPOLOGY: linear
45 MOLECULE TYPE: DNA (genomic)
46 HYPOTHETICAL: NO
47 ANTI-SENSE: NO
48
49 ORIGINAL SOURCE:
50 ORGANISM: Trichoderma reesei
51 FEATURE:
52 NAME/KEY: CDS
53 LOCATION: join(311..375, 446..2205, 2270..2675)
54 FEATURE:
55 NAME/KEY: intron
56 LOCATION: 376..445
57 FEATURE:
58 NAME/KEY: intron
59 LOCATION: 2206..2269
60
61 15-08-08-462-090-1

```

Query Match	20.8%	score 25.6;	DB 3;	Length 3033;-
Best Local Similarity	55.7%;	pred. No. 3.9;		
Matches 49:	Conservative	0;	Mismatches	39;
			Indels	0;
			Gaps	0;

12 AGTTTGTCTAGTGTCGAAAGGCCCAACACATGTGTTCTTGCCAGTAGTTAGGTGAC 71
||||| |||| |
Db AGTTTCTCCTACAGCTCTTGACCAAGACCATTCTGTGAGCCCAATCAGAAATCGGTTAC 319

72 AGAARGGGGCTTTAGCACTAGCGCTTGACA 99
||||| |||| |
Db CGAARGCAGCAGTGCCTGCTGCACTTGCCA 347

RESULT 4
US-08-463-461-1
; Sequence 1, Application US/08463461
; Patent No. 6103464

```

GENERAL INFORMATION:
APPLICANT: Fowler, Timothy
APPLICANT: Barnett, Christopher C.
APPLICANT: Shoemaker, Sharon
TITLE OF INVENTION: Saccharification of Cellulose by Cloning
TITLE OF INVENTION: and Amplification of the Beta-glucosidase Gene of
TITLE OF INVENTION: Trichoderma Reesei
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Genencor International, Inc.
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: U.S.A.
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,461
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Christopher L. Stone
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC78D4
TELEPHONE: 415-846-7555
TELEFAX: 415-845-6504
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3033 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoderma reesei
FEATURE:
NAME/KEY: CDS
LOCATION: join(311..375, 446..2205, 2270..2679)
FEATURE:
NAME/KEY: intron
LOCATION: 376..445
FEATURE:
NAME/KEY: intron
LOCATION: 2206..2269
US-08-463-461-1

Query Match 20.8%; Score 25.6; DB 3; Length 3033;
Best Local Similarity 55.7%; Pred. No. 3.9; Mismatches 0; Gaps 0;
Matches 49; Conservative 0; Indels 39;

QY 12 AGTTGTTGTTAGTGCAGAGGCCCAACACTGTGTTCTTCCAGTAGTGGTTGTAC 71
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 260 AGTTCTCTACCACTCTTGACCAAGACCATTCTGTGAGCCCAATCAGAAATGGTTAC 319
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 72 AGAAGCGGTTAGCACTAGCGCTTGACA 99
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 320 CGAAGCAGCTGCGCTGGCACTTGCCA 347
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
US-08-614-770A-1/C
; Sequence 1, Application US/08614770A
; Patent No. 5773267
; GENERAL INFORMATION:
; APPLICANT: WILLIAM R. JACOBS AND GRAHAM F. HATFULL
; TITLE OF INVENTION: D29 SHUTTLE PHASMIDS AND USES THEREOF

```

```

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESS: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
MEDIUM TYPE: DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/614,770A
FILING DATE: MARCH 7, 1996
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 96700/402
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 49272
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: OLIGONUCLEOTIDE
DESCRIPTION: NO
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: MYCOBACTERIOPHAGE
INDIVIDUAL ISOLATE: D29
US-08-614-770A-1

Query Match 20.5%; Score 25.2; DB 1; Length 49272;
Best Local Similarity 55.8%; Pred. No. 16;
Matches 48; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 22 AGTGTGGAAGGCCCAACACTGTGTTCTTCCAGTAGTGGTTGTACAGAGCGCT 81
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 47920 AGTGTCAACGTCCGACCTCCGAGATCGTCGACAGCGTGAGCTGCCCTGTGCGAAGA 47861
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 82 TAGCACTAGCGCTTGACAGAACCTCA 107
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 47860 CATCAGATCACTGGCAGACCGTCA 47835
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
US-08-705-771-1
; Sequence 1, Application US/08705771
; Patent No. 6054289
; GENERAL INFORMATION:
; APPLICANT: Paul Moore, Reiner Gentz, Hongjin Ji,
; APPLICANT: Jian Ni and Jing-Shan Hu
; TITLE OF INVENTION: Human Genes, Sequences and
; TITLE OF INVENTION: Expression Products
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE

```


Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCNDA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 133
LENGTH: 588
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(588)
OTHER INFORMATION: n - A,T,C or G
US-09-385-982-133

Query Match 19.0%; Score 23.4; DB 4; Length 588;
Best Local Similarity 92.3%; Pred. No. 14;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 97 ACAGACCTTCACAGACCCCAAGGTAC 122
DB 30 ACAGACCTTCACAGACCCCAAGGAC 55

RESULT 13
US-07-927-391-25/C
Sequence 25, Application US/07927391
Patent No. 6001649
GENERAL INFORMATION:
APPLICANT: CAPUT, Daniel
APPLICANT: FERRARA, Pascual
APPLICANT: MILOUX, Brigitte
APPLICANT: MINTY, Adrian
APPLICANT: VITA, Natalio
TITLE OF INVENTION: Protein having a cytokin type
TITLE OF INVENTION: activity and recombinant DNA, expression vector and hosts
TITLE OF INVENTION: for its preparation.
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: King Street Station, Suite 500, 1800 Diagonal
STREET: Road, PO Box 299
CITY: ALEXANDRIA
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/927,391
FILING DATE: 19920929
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: SAXE, Bernhard D
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 16781/369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109

Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCNDA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 133
LENGTH: 588
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(588)
OTHER INFORMATION: n - A,T,C or G
US-09-385-982-133

Query Match 19.0%; Score 23.4; DB 4; Length 588;
Best Local Similarity 92.3%; Pred. No. 14;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 97 ACAGACCTTCACAGACCCCAAGGTAC 122
DB 30 ACAGACCTTCACAGACCCCAAGGAC 55

RESULT 13
US-07-927-391-25/C
Sequence 25, Application US/07927391
Patent No. 6001649
GENERAL INFORMATION:
APPLICANT: CAPUT, Daniel
APPLICANT: FERRARA, Pascual
APPLICANT: MILOUX, Brigitte
APPLICANT: MINTY, Adrian
APPLICANT: VITA, Natalio
TITLE OF INVENTION: Protein having a cytokin type
TITLE OF INVENTION: activity and recombinant DNA, expression vector and hosts
TITLE OF INVENTION: for its preparation.
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: King Street Station, Suite 500, 1800 Diagonal
STREET: Road, PO Box 299
CITY: ALEXANDRIA
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/927,391
FILING DATE: 19920929
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: SAXE, Bernhard D
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 16781/369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109

Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCNDA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 133
LENGTH: 588
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(588)
OTHER INFORMATION: n - A,T,C or G
US-09-385-982-133

Query Match 19.0%; Score 23.4; DB 4; Length 588;
Best Local Similarity 92.3%; Pred. No. 14;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 97 ACAGACCTTCACAGACCCCAAGGTAC 122
DB 30 ACAGACCTTCACAGACCCCAAGGAC 55

RESULT 13
US-07-927-391-25/C
Sequence 25, Application US/07927391
Patent No. 6001649
GENERAL INFORMATION:
APPLICANT: CAPUT, Daniel
APPLICANT: FERRARA, Pascual
APPLICANT: MILOUX, Brigitte
APPLICANT: MINTY, Adrian
APPLICANT: VITA, Natalio
TITLE OF INVENTION: Protein having a cytokin type
TITLE OF INVENTION: activity and recombinant DNA, expression vector and hosts
TITLE OF INVENTION: for its preparation.
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: King Street Station, Suite 500, 1800 Diagonal
STREET: Road, PO Box 299
CITY: ALEXANDRIA
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/927,391
FILING DATE: 19920929
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: SAXE, Bernhard D
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 16781/369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109

TELEX: 899149
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 741 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-07-927-391-25

Query Match 19.0%; Score 23.4; DB 3; Length 741;
Best Local Similarity 60.0%; Pred. No. 15;
Matches 39; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 13 GTTTGTTCTAGTGTGCAAGAGCCCACTGTGTTCTTCCCACTGAGTTAGTTGTACA 72
DB 231 GATCTCTATAGTCTGCGAGCCTCTGCACTGAGATCTTCTATTTGTTGGAATATAACA 172

QY 73 GAACG 77
DB 171 GCAGG 167

RESULT 14
US-08-086-428B-11/c
Sequence 11, Application US/08086428B
Patent No. 5514539
GENERAL INFORMATION:
APPLICANT: BURK, J., MILLER, R.H. AND
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE
TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE
TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN
NUMBER OF SEQUENCES: 159
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/086,428B
FILING DATE: 29-JUN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4070
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 576 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: DK1
US-08-086-428B-11

Query Match 18.9%; Score 23.2; DB 1; Length 576;
Best Local Similarity 61.7%; Pred. No. 16;

Matches 37; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 35 CCCAACACTGTGTTCTTCCAGTCAGTTAGTGTGTACAGAACGGCGTTAGCACTAGCGCT 94
 III III III I III I III III III III III III III III
 Db 476 CCCGCCACCATGTCACGACAGCTTGTGGGATTCGGAGTAACTCGGATAGCAGCTAGGGCT 417

RESULT 15

US-08-468-570-11/c
 ; Sequence 11, Application US/08468570
 ; Patent No. 5871962
 ; GENERAL INFORMATION:
 ; APPLICANT: BURK, J., MILLER, R.H. AND
 ; APPLICANT: PORCELL, R.H.
 ; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 ; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE
 ; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE
 ; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN
 ; TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES
 ; NUMBER OF SEQUENCES: 159
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/468,570
 ; FILING DATE: 6-JUN-1995
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/086,428
 ; FILING DATE: 29-JUN-1993
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: RICHARD W. BORK
 ; REGISTRATION NUMBER: 36,459
 ; REFERENCE/DOCKET NUMBER: 2026-4070051
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ; TELEX: 421792

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:
 LENGTH: 576 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 ORIGINAL SOURCE:
 ORGANISM: homoplans
 INDIVIDUAL ISOLATE: DK1
 US-08-468-570-11

Query Match 18.9%; Score 23.2; DB 2; Length 576;
 Best Local Similarity 61.7%; Pred. No. 16;
 Matches 37; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 35 CCCAACACTGTGTTCTTCCAGTCAGTTAGTGTGTACAGAACGGCGTTAGCACTAGCGCT 94
 III III III I III I III III III III III III III III
 Db 476 CCCGCCACCATGTCACGACAGCTTGTGGGATTCGGAGTAACTCGGATAGCAGCTAGGGCT 417

Search completed: October 26, 2002, 04:59:35
 Job time : 54 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	123	100.0	123	22	AAH22395	Human rac1 genomic
C 2	123	100.0	348	22	AAH22399	Human rac1 gene re
3	123	100.0	1266	22	AAH22396	Human rac1 contig
C 4	106.6	86.7	422	21	AAH9694	Mouse Rab2 nucleot
C 5	96.6	78.5	447	21	AAH89693	Mouse Exol03 nucle
C 6	79.6	64.7	352	22	AAH22400	Human rac1 gene re
C 7	79.6	64.7	1318	23	AA655569	DNA encoding novel
C 8	79.6	64.7	2051	21	AAF181236	Lung cancer associ
C 9	38.2	31.1	3740	22	AAI07296	Human reproductive

PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels

PS Disclosure; Page 7; 76pp; English.

XX The present invention describes a method of screening a test agent for
XX hepatotoxicity. The method comprises: (a) providing a test cell
XX population comprising a cell capable of expressing one or more nucleic
XX acid sequences selected from the group consisting of RISKMARKER 1-8
XX and INJURYMARKER 1-10; (b) contacting the test cell population with a
XX test agent; (c) measuring expression of one or more of the nucleic
XX acid sequences in the test cell population; (d) comparing the
XX expression of the nucleic acid sequence in the test cell population to
XX the expression of the nucleic acid sequence in a reference cell
XX population comprising at least one cell whose exposure status to a
XX hepatotoxic agent is known; and (e) identifying a difference in
XX expression levels of the RISKMARKER or INJURYMARKER sequences, if
XX present, in the test cell population and reference cell population.
XX The method is useful for identifying a hepatotoxic agent. The present
XX sequence is given in the exemplification of the present invention.

SQ Sequence 123 BP; 36 A; 28 C; 30 G; 29 T; 0 other;

Query Match 100.0%; Score 123; DB 22; Length 123;
Best Local Similarity 100.0%; Pred. No. 8.6e-36;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAGTTTCTAGTGTGCGAAAGGCCCAACACATGTCTTTCAGTGAG 60
|||||

Db 1 CAATTGAAAAGTTTCTAGTGTGCGAAAGGCCCAACACATGTCTTTCAGTGAG 60
|||||

QY 61 TTAGTGTGTACAGACGGCTTAGCCTAGCGCTTGACAGACCTCACAGACCCAAAGT 120
|||||

Db 61 TTAGTGTGTACAGACGGCTTAGCCTAGCGCTTGACAGACCTCACAGACCCAAAGT 120
|||||

QY 121 ACC 123
|||

Db 121 ACC-123

RESULT 2

AAH22399/c

ID AAH22399 standard; DNA; 348 BP.

AC AAH22399;

XX 22-AUG-2001 (first entry)

DT Human rac1 gene related nucleotide sequence #3.

XX Identification; toxic; hepatotoxic; differential gene expression;

KW NSAID; non-steroidal antiinflammatory drug; ds.

XX Homo sapiens.

XX WO200138579-A2.

XX 31-MAY-2001.

XX 21-NOV-2000; 2000WO-US32049.

XX 22-NOV-1999; 99US-0166923.

PR 18-FEB-2000; 2000PS-0183531.

PR 20-NOV-2000; 2000US-0717321.

XX (CURA-) CURAGEN CORP.

XX Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;

XX WPI; 2001-355948/37.

PT Screening hepatotoxic agent comprises contacting test cell population

PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels

PS Disclosure; Page 8-9; 76pp; English.

XX The present invention describes a method of screening a test agent for
XX hepatotoxicity. The method comprises: (a) providing a test cell
XX population comprising a cell capable of expressing one or more nucleic
XX acid sequences selected from the group consisting of RISKMARKER 1-8
XX and INJURYMARKER 1-10; (b) contacting the test cell population with a
XX test agent; (c) measuring expression of one or more of the nucleic
XX acid sequences in the test cell population; (d) comparing the
XX expression of the nucleic acid sequence in the test cell population to
XX the expression of the nucleic acid sequence in a reference cell
XX population comprising at least one cell whose exposure status to a
XX hepatotoxic agent is known; and (e) identifying a difference in
XX expression levels of the RISKMARKER or INJURYMARKER sequences, if
XX present, in the test cell population and reference cell population.
XX The method is useful for identifying a hepatotoxic agent. The present
XX sequence is given in the exemplification of the present invention.

SQ Sequence 348 BP; 103 A; 76 C; 67 G; 102 T; 0 other;

Query Match 100.0%; Score 123; DB 22; Length 348;
Best Local Similarity 100.0%; Pred. No. 1.3e-35;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAGTTTCTAGTGTGCGAAAGGCCCAACACATGTCTTTCAGTGAG 60
|||||

Db 212 CAATTGAAAAGTTTCTAGTGTGCGAAAGGCCCAACACATGTCTTTCAGTGAG 153
|||||

QY 61 TTAGTGTGTACAGACGGCTTAGCCTAGCGCTTGACAGACCTCACAGACCCAAAGT 120
|||||

Db 152 TTAGTGTGTACAGACGGCTTAGCCTAGCGCTTGACAGACCTCACAGACCCAAAGT 93
|||||

QY 121 ACC 123
|||

Db 92 ACC 90

RESULT 3

AAH22396

ID AAH22396 standard; DNA; 1266 BP.

AC AAH22396;

XX 22-AUG-2001 (first entry)

DT Human rac1 contig SEQ ID NO:2.

XX Identification; toxic; hepatotoxic; differential gene expression;

KW NSAID; non-steroidal antiinflammatory drug; ds.

XX Homo sapiens.

XX WO200138579-A2.

XX 31-MAY-2001.

XX 21-NOV-2000; 2000WO-US32049.

XX 22-NOV-1999; 99US-0166923.

PR 18-FEB-2000; 2000PS-0183531.

PR 20-NOV-2000; 2000US-0717321.

XX (CURA-) CURAGEN CORP.

XX Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;

XX WPI; 2001-355948/37.

PT Screening hepatotoxic agent comprises contacting test cell population

expressing RISKMARKER or INJURYMARKER with agent, comparing expression levels with reference population and identifying difference in expression

Disclosure; Page 7; 76pp; English.

The present invention describes a method of screening a test agent for hepatotoxicity. The method comprises: (a) providing a test cell population comprising a cell capable of expressing one or more nucleic acid sequences selected from the group consisting of RISKMARKER 1-8 and INJURYMARKER 1-10; (b) contacting the test cell population with a test agent; (c) measuring expression of one or more of the nucleic acid sequences in the test cell population; (d) comparing the expression of the nucleic acid sequence in the test cell population to the expression of the nucleic acid sequence in an reference cell population comprising at least one cell whose exposure status to a hepatotoxic agent is known; and (e) identifying a difference in expression levels of the RISKMARKER or INJURYMARKER sequences, if present, in the test cell population and reference cell population. The method is useful for identifying a hepatotoxic agent. The present invention is given in the exemplification of the present invention.

Sequence 1266 BP; 385 A; 258 C; 285 G; 338 T; 0 other;

Query Match 100.0%; Score 123; DB 22; Length 1266;

Best Local Similarity 100.0%; Pred. No. 2.1e-35;

Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAGTTTCTAGTGTGCGAAGGCCCAACACATGTTCTTCCAGTGAG 60

|||||

Db 146 CAATTGAAAAGTTTCTAGTGTGCGAAGGCCCAACACATGTTCTTCCAGTGAG 205

|||||

QY 61 TTAGTTGTACAGAACGGCGTTAGCTAGCGTTGACAGAACCTCACAGACCCAAAGGT 120

|||||

Db 206 TTAGTTGTACAGAACGGCGTTAGCTAGCGTTGACAGAACCTCACAGACCCAAAGGT 265

|||||

QY 121 ACC 123

|||

Db 266 ACC 268

RESULT 4

AAA89694/c

ID AAA89694 standard; cDNA; 422 BP.

XX AC

AAA89694;

XX DT

08-JAN-2001 (first entry)

XX

Mouse Rab2 nucleotide sequence #1.

XX

Exocytosis pathway protein; Exo protein; antiparkinsonian; antidiabetic; antiallergic; antiasthmatic; nootropic; neuroprotective; anticonvulsant; vulnarary; asthma; inflammation; allergy; Chediak-Higashi syndrome; CHS; Alzheimer's disease; Parkinson's disease; Huntington's disease; diabetes; digestion disorder; wound healing disorder; gene therapy; ss.

XX OS

XX Mus sp.

XX

PN WO200043419-A2.

XX

PD 27-JUL-2000.

XX

XX 20-JAN-2000; 2000WO-US01431.

XX

PR 20-JAN-1999; 99US-0116534.

XX

PR 26-JAN-1999; 99US-0117274.

XX

PR 26-JAN-1999; 99US-0117308.

XX

PR 26-JAN-1999; 99US-0117309.

XX

PR 26-JAN-1999; 99US-0117312.

XX

PR 01-FEB-1999; 99US-0118177.

XX

PR 01-FEB-1999; 99US-0118178.

XX

PR 01-FEB-1999; 99US-0118179.

XX

PR 09-FEB-1999; 99US-0119286.

PR 11-FEB-1999; 99US-0119998.

PR 11-FEB-1999; 99US-0119759.

XX

PA (RIGE-) RIGEL PHARM INC.

XX

PI Luo Y;

XX

XX WPI; 2000-482908/42.

DR

XX New nucleic acids encoding Exo proteins which are useful in the diagnosis, treatment or prevention of exocytosis-mediated disorders such as asthma, inflammation and allergies -

PT

PT

XX

PS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PF 20-JAN-2000; 2000WO-US01431.
XX 20-JAN-1999; 99US-0116534.
PR 26-JAN-1999; 99US-0117274.
PR 26-JAN-1999; 99US-0117308.
PR 26-JAN-1999; 99US-0117309.
PR 26-JAN-1999; 99US-0117312.
PR 01-FEB-1999; 99US-0118177.
PR 01-FEB-1999; 99US-0118178.
PR 01-FEB-1999; 99US-0118179.
PR 09-FEB-1999; 99US-0119286.
PR 11-FEB-1999; 99US-0119998.
PR 11-FEB-1999; 99US-0119759.
XX (RIGE-) RIGEL PHARM INC.
XX
XX Luo Y;
PI
XX
XX WPI; 2000-482908/42.
DR
XX
XX New nucleic acids encoding Exo proteins which are useful in the
PT diagnosis, treatment or prevention of exocytosis-mediated disorders
PT such as asthma, inflammation and allergies -
XX
XX
PS Disclosure; Page 271; 305pp; English.
XX
XX The present sequence encodes a polypeptide which is associated with
CC the exocytosis pathway. cDNA molecules encoding proteins involved in
CC exocytosis have been isolated by yeast one-hybrid and two-hybrid
CC screening. Novel proteins, termed Exo proteins, have been identified that
CC interact with known exocytosis-associated proteins such as GS27, alpha
CC snap, unc18-1, vamps3, snap23, and the rab family of proteins.
CC Exo proteins and their agonists and antagonists are useful in the
CC diagnosis, treatment or prevention of exocytosis-mediated disorders
CC such as asthma, inflammation, allergies, Chediak-Higashi Syndrome
CC (CHS), Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC diabetes, digestion disorders and wound healing disorders.
CC The nucleic acids, antagonists or agonists of Exo proteins are useful
CC in gene therapy. The nucleic acids are also useful for generating
CC transgenic or knock-out animals which can be used in the
CC development and screening of therapeutically useful reagents.
XX
XX Sequence 447 BP; 102 A; 95 C; 93 G; 157 T; 0 other;
SQ
Query Match 78.5%; Score 96.6; DB 21; Length 447;
Best Local Similarity 91.9%; Pred. No. 8.3e-26;
Matches 113; Conservative 0; Mismatches 9; Indels 1; Gaps 1;
QY 1 CAATTGAAAAAGTTTGTTC-TAGTGGTGAAGGCCCAACACTGTGTTTCCAGTGA 59
DB 379 CAATTGAAAAAGTTTGTTC-TAGTGGTGAAGGCCCAACACTGTGTTTCCAGTGA 320
QY 60 GTTAGGTTGTACAGACCGCGTTAGCACTAGCGCTTGACAGACCTCAGACCCCAAGG 119
DB 319 GTTAGGTTGTACAGACCGCGTTAGCACTAGCGCTTGACAGACCTCAGACCCCAAGG 260
QY 120 TAC 122
DB 259 AAC 257
RESULT 6
AAH22400/c
ID AAH22400 standard; DNA; 352 BP.
XX
AC AAH22400;
XX
XX 22-AUG-2001 (first entry)
XX
XX Human rac1 gene related nucleotide sequence #4.
XX
XX Identification; toxic; hepatotoxic; differential gene expression;
KW NSAID; non-steroidal antiinflammatory drug; ds.
XX

XX Homo sapiens.
XX
XX WO200138579-A2.
XX
XX 31-MAY-2001.
XX
XX 21-NOV-2000; 2000WO-US32049.
XX
XX 22-NOV-1999; 99US-0166923.
PR 18-FEB-2000; 2000US-0183531.
PR 20-NOV-2000; 2000US-0717321.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Gould-Rothberg BE, Dipippo VA, Ramsen TM, Gerwein RW;
XX
XX WPI; 2001-355948/37.
XX
XX Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels -
XX
XX Disclosure; Page 8-9; 76pp; English.
XX
XX The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in a reference cell
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent is known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
XX
XX Sequence 352 BP; 86 A; 77 C; 72 G; 117 T; 0 other;
SQ
Query Match 64.7%; Score 79.6; DB 22; Length 352;
Best Local Similarity 83.1%; Pred. No. 1.5e-19;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;
QY 1 CAATTGAAAAAGTTTGTTC-TAGTGGTGAAGGCCCAACACTGTGTTTCCAGTGA 60
DB 216 CAATTGAAAAAGTTTGTTC-TAGTGGTGAAGGCCCAACACTGTGTTTCCAGTGA 157
QY 61 TTAGGTTGTACAGACCGCGTTAGCACTAGCGCTTGACAGACCTCAGACCCCAAG 118
DB 156 TTAGGTTGTACAGACCATCGTCAGCACTAGCACTAGCACTAGCACTAGCACTAG 97
QY 119 GTAC 122
DB 96 GAC 93
RESULT 7
AAS65569/c
ID AAS65569 standard; cDNA; 1318 BP.
XX
AC AAS65569;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #1373.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX

```
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang VT;
XX PD WPI; 2001-639362/73.
XX PR P-PSDB; ABO1382.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 1; SEQ ID No 1373; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94584 represent novel human
XX CC diagnostic coding sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX CC Sequence 1318 BP; 346 A; 268 C; 279 G; 424 T; 1 other;
XX CC
XX CC Query Match 64.7%; Score 79.6; DB 23; Length 1318;
XX CC Best Local Similarity 83.1%; Pred. No. 2.4e-19;
XX CC Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;
XX CC
QY 1 CAATTGAAAAGTTGTTCTAGTGGTGAAGAGGCCCAACACTGTGTTCTGCCAGTGAG 60
DB 1221 CAATTTAAAAAATTCGTGTTAGTGGCTGAAGGTCGCCAGCTGTATTCGCCAGTGAG 1162
QY 61 TTAGGTGTACAGACGGCTTACCACTAGCGC--TTGACAGAACCTCAGACACCCAAAG 118
DB 1161 TTAGTGTACAGAACATCGTCAGCACTAGCAGATTACAGAACTTCACAGACCCAAAG 1102
QY 119 GTAC 122
DB 1101 GAAC 1098
XX CC
RESULT 8
ID AAF18236/c
XX ID AAF18236 standard; DNA; 2051 BP.
XX AC AAF18236;
XX AC
```

```
DT XX 14-MAR-2001 (first entry)
DE XX Lung cancer associated polynucleotide sequence SEQ ID 255.
KW XX Human; lung cancer associated protein; neuroprotective; cytostatic;
KW XX cardioactive; immunomodulatory; muscular active; vulnerary;
KW XX gastrointestinal; nephrotropic; antineoplastic; gynecological;
KW XX antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW XX proliferative disorder; wound healing; infectious disease; ds.
XX OS Homo sapiens.
XX PN WO200055180-A2.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US05918.
XX PR 12-MAR-1999; 99US-0124270.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (ROSE/) ROSEN C A.
XX PI Ruben SM;
XX DR WPI; 2000-587514/55.
XX DR P-PSDB; AAS58360.
XX PT Lung cancer associated gene sequences, referred to as lung cancer
XX PT antigens, useful for treatment, prevention, and diagnosis of disorders
XX PT such as lung cancer.
XX PS Claim 1; Page 716-717; 1425pp; English.
XX CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
XX CC associated proteins represented in AAS58106 - AAS58548. Lung cancer
XX CC associated proteins and polynucleotide sequences, their agonists, and
XX CC antagonists may have neuroprotective; cytostatic; cardioactive;
XX CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
XX CC general; nephrotropic; antineoplastic; gynecological; or antibacterial
XX CC activity. The invention also includes antibodies specific for the
XX CC protein or polynucleotide sequences. The lung cancer associated
XX CC polynucleotide sequences may be used for detection of lung cancer,
XX CC chromosome identification, as chromosome markers, and for numerous other
XX CC diagnostic or research purposes. The proteins may be used to treat
XX CC disorders such as neural, immune, muscular, reproductive,
XX CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX CC disorders. The proteins may also be used in the treatment of wounds and
XX CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
XX CC peptide AAS5549 are used in the course of the invention for the
XX CC identification and characterisation of the polynucleotide and protein
XX CC sequences.
XX CC Sequence 2051 BP; 570 A; 430 C; 433 G; 612 T; 6 other;
XX CC
XX CC Query Match 64.7%; Score 79.6; DB 21; Length 2051;
XX CC Best Local Similarity 83.1%; Pred. No. 2.9e-19;
XX CC Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;
XX CC
QY 1 CAATTGAAAAGTTGTTCTAGTGGTGAAGAGGCCCAACACTGTGTTCTGCCAGTGAG 60
DB 1896 CAATTTAAAAAATTCGTGTTAGTGGCTGAAGGTCGCCAGCTGTATTCGCCAGTGAG 1837
QY 61 TTAGGTGTACAGACGGCTTACCACTAGCGC--TTGACAGAACCTCAGACACCCAAAG 118
DB 1836 TTAGTGTACAGAACATCGTCAGCACTAGCAGATTACAGAACTTCACAGACCCAAAG 1777
QY 119 GTAC 122
DB 1776 GAAC 1773
XX CC
RESULT 9
```

AA07296/c
ID AAL07296 standard; DNA; 3740 BP.
XX
AC AAL07296;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 9984.
DE Human reproductive system related antigen; reproductive system disorder;
KW Human; reproductive system related antigen; cancer; gene therapy; ds.
KW
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
31-JAN-2000; 2000US-0179065.
04-FEB-2000; 2000US-0180628.
24-FEB-2000; 2000US-0184564.
02-MAR-2000; 2000US-0186350.
16-MAR-2000; 2000US-0189874.
17-MAR-2000; 2000US-0190076.
18-APR-2000; 2000US-0198123.
19-MAY-2000; 2000US-0205515.
07-JUN-2000; 2000US-0209467.
28-JUN-2000; 2000US-0214886.
30-JUN-2000; 2000US-0215135.
07-JUL-2000; 2000US-0216647.
07-JUL-2000; 2000US-0216880.
11-JUL-2000; 2000US-0217487.
14-JUL-2000; 2000US-0217496.
14-JUL-2000; 2000US-0218290.
26-JUL-2000; 2000US-0220963.
26-JUL-2000; 2000US-0220964.
14-AUG-2000; 2000US-0224518.
14-AUG-2000; 2000US-0224519.
14-AUG-2000; 2000US-0225213.
14-AUG-2000; 2000US-0225214.
14-AUG-2000; 2000US-0225266.
14-AUG-2000; 2000US-0225267.
14-AUG-2000; 2000US-0225268.
14-AUG-2000; 2000US-0225270.
14-AUG-2000; 2000US-0225447.
14-AUG-2000; 2000US-0225757.
14-AUG-2000; 2000US-0225758.
14-AUG-2000; 2000US-0225759.
18-AUG-2000; 2000US-0226279.
22-AUG-2000; 2000US-0226681.
22-AUG-2000; 2000US-0226868.
22-AUG-2000; 2000US-0227182.
23-AUG-2000; 2000US-0227009.
30-AUG-2000; 2000US-0228924.
01-SEP-2000; 2000US-0229287.
01-SEP-2000; 2000US-0229343.
01-SEP-2000; 2000US-0229344.
01-SEP-2000; 2000US-0229345.
05-SEP-2000; 2000US-0229509.
05-SEP-2000; 2000US-0229513.
06-SEP-2000; 2000US-0230437.
06-SEP-2000; 2000US-0230438.
08-SEP-2000; 2000US-0231242.
08-SEP-2000; 2000US-0231243.
08-SEP-2000; 2000US-0231244.
08-SEP-2000; 2000US-0231413.
08-SEP-2000; 2000US-0231414.
08-SEP-2000; 2000US-0232080.
08-SEP-2000; 2000US-0232081.
12-SEP-2000; 2000US-0231968.
14-SEP-2000; 2000US-0232397.
14-SEP-2000; 2000US-0232398.
14-SEP-2000; 2000US-0232399.
14-SEP-2000; 2000US-0232400.
14-SEP-2000; 2000US-0232401.
14-SEP-2000; 2000US-0233063.
14-SEP-2000; 2000US-0233064.
14-SEP-2000; 2000US-0233065.
21-SEP-2000; 2000US-0234223.
21-SEP-2000; 2000US-0234274.
25-SEP-2000; 2000US-0234997.
25-SEP-2000; 2000US-0234998.
26-SEP-2000; 2000US-0234984.
27-SEP-2000; 2000US-0235834.
27-SEP-2000; 2000US-0235836.
29-SEP-2000; 2000US-0236327.
29-SEP-2000; 2000US-0236367.
29-SEP-2000; 2000US-0236368.
29-SEP-2000; 2000US-0236369.
29-SEP-2000; 2000US-0236370.
02-OCT-2000; 2000US-0236802.
02-OCT-2000; 2000US-0237037.
02-OCT-2000; 2000US-0237038.
02-OCT-2000; 2000US-0237039.
13-OCT-2000; 2000US-0239935.
13-OCT-2000; 2000US-0239937.
20-OCT-2000; 2000US-0240960.
20-OCT-2000; 2000US-0241221.
20-OCT-2000; 2000US-0241785.
20-OCT-2000; 2000US-0241786.
20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
01-NOV-2000; 2000US-0244617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
17-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.
17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249264.
17-NOV-2000; 2000US-0249265.
17-NOV-2000; 2000US-0249297.
17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.

05-DEC-2000; 2000US-0251988.
06-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition -
XX Disclosure; SEQ ID NO 984; 1297pp + Sequence Listing; English.
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention.
XX Sequence 3740 BP; 1188 A; 752 C; 687 G; 1113 T; 0 other;
Query Match 31.1%; Score 38.2; DB 22; Length 3740;
Best Local Similarity 78.0%; Pred. No. 0.00076;
Matches 46; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 8 AAAAGTTCTCTAGTGGTGAAGGCCCAACACTGTTCTTGCAGTACGTAGGT 66
DB 2725 AAAATATTGTTCTAGTGTGTGAAGGCCCAAGTGTATATCTTGCAGTACGTAAAGT 2667
RESULT 10
AA16478
ID AA16478 standard; DNA; 605 BP.
XX AC
XX AA16478;
XX 14-JUN-2000 (first entry)
XX Human colon cancer differentially expressed nucleotide sequence #483.
XX Colon cancer; detect; differential expression; human; treatment;
XX detect mutation; non-invasive diagnostic method; ds..
XX Homo sapiens.
XX WO200012702-A2.
XX 09-MAR-2000.
XX 30-AUG-1999; 99WO-US19424.
XX 31-AUG-1998; 98US-0098639.
XX 27-JAN-1999; 99US-0117393.
XX (FARB) BAYER CORP.
XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
XX Catino TJ, Dwivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;
XX Schlegel R;
XX WPI; 2000-256641/22.
XX Novel nucleic acids and proteins for identifying therapeutic agents
XX useful for treating and diagnosing cancer, especially colon cancer -
PT

Claim 16; Page 319; 345pp; English.
XX This sequence represents a human nucleotide sequence which is
XX differentially expressed in colon cancer cells compared to the expression
XX levels in normal cells. The nucleotide sequence can be used as a source
XX of primers and probes. The nucleotide sequence is useful for determining
XX the phenotype of a cell by detecting the differential expression of the
XX sequence relative to a normal cell. The probes derived from the sequence
XX can also be used to determine the phenotype of cells in a sample. Probes
XX and antibodies which hybridise to the nucleotide sequence can also be
XX used to determine the phenotype of a cell. The primers are useful for
XX detecting a mutation in a test nucleotide sequence and also for detecting
XX cancer, preferably colon cancer. Antibodies against the protein encoded
XX by the nucleotide sequence can also be used in a method to detect colon
XX cancer. The diagnostic method is non-invasive and accurate for diagnosing
XX colon cancer at an early stage.
XX Sequence 605 BP; 214 A; 105 C; 116 G; 155 T; 15 other;
Query Match 25.5%; Score 31.4; DB 21; Length 605;
Best Local Similarity 85.5%; Pred. No. 0.13;
Matches 47; Conservative 0; Mismatches 6; Indels 2; Gaps 1;
QY 70 ACAGAACGGCTTAGCAGTACGCGC--TTGACAGAACCTCAGACGCCCAAGGTAC 122
DB 1 ACAGAACATCGTCAGCAGTACGAGTTTACAGAACCTCAGACGCCCAAGGTAC 55
RESULT 11
AAC50142
ID AAC50142 standard; DNA; 1089 BP.
XX AC
XX AAC50142;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 63742.
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 21-APR-1999; 99US-0130077.
XX 19-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.

PR	11-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	03-AUG-1999;	99US-0147038.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147204.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147302.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147192.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147260.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147303.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147416.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139452.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140891.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141187.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159283.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	22-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.

PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match 22.3%; Score 27.4; DB 21; Length 1089;			
Best Local Similarity 57.6%; Pred. No. 4.7;			
Matches 49; Conservative 0; Mismatches 36; Indels 0; Gaps 0;			
QY	23	GTGTCGAAGGCCACACATGTTCTTCCAGTCAGTTAGTTGTACAGACGGCGTT	82
DB	699	GAGATTGAAACAAACACGAGATACATCGCTCTTTATAAGTTGGAGAGATGGGAT	758
QY	83	AGCACTAGCGCTTGACAGAACCTCA	107
DB	759	ATCGCTATGCTGCACCTACCTCA	783
RESULT 12			
AAC40517			
ID	AAC40517 standard; DNA; 1091 BP.		
XX	AAC40517;		
AC	AAC40517;		
XX	17-OCT-2000 (first entry)		
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 28592.		
XX	Hybridisation assay; genetic mapping; gene expression control;		
KW	protein identification; signal transduction pathway;		
KW	metabolic pathway; promoter; termination sequence; ss.		
XX	Arabidopsis thaliana.		
XX	EF1033405-A2.		
XX	06-SEP-2000.		
XX	25-FEB-2000; 2000EP-0301439.		
XX	25-FEB-1999; 99US-0121825.		
XX	05-MAR-1999; 99US-0123180.		
XX	09-MAR-1999; 99US-0123548.		
XX	23-MAR-1999; 99US-0125788.		
XX	25-MAR-1999; 99US-0126264.		
XX	29-MAR-1999; 99US-0126785.		
XX	01-APR-1999; 99US-0127462.		
XX	06-APR-1999; 99US-0128234.		
XX	08-APR-1999; 99US-0128714.		
XX	16-APR-1999; 99US-0129845.		
XX	19-APR-1999; 99US-0130077.		
XX	21-APR-1999; 99US-0130449.		
XX	23-APR-1999; 99US-0130510.		
XX	23-APR-1999; 99US-0130891.		
XX	28-APR-1999; 99US-0131449.		
XX	30-APR-1999; 99US-0132048.		
XX	04-MAY-1999; 99US-0132407.		
XX	05-MAY-1999; 99US-0132484.		
XX	06-MAY-1999; 99US-0132485.		
XX	06-MAY-1999; 99US-0132486.		
XX	07-MAY-1999; 99US-0132863.		
XX	11-MAY-1999; 99US-0132863.		
XX	14-MAY-1999; 99US-0134256.		
XX	14-MAY-1999; 99US-0134218.		
XX	14-MAY-1999; 99US-0134219.		
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135829.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136782.	
PR	01-JUN-1999;	99US-0137222.	
PR	03-JUN-1999;	99US-0137528.	
PR	04-JUN-1999;	99US-0137502.	
PR	07-JUN-1999;	99US-0137724.	
PR	08-JUN-1999;	99US-0138094.	
PR	10-JUN-1999;	99US-0138540.	
PR	10-JUN-1999;	99US-0138847.	
PR	14-JUN-1999;	99US-0139119.	
PR	16-JUN-1999;	99US-0139452.	
PR	16-JUN-1999;	99US-0139453.	
PR	17-JUN-1999;	99US-0139492.	
PR	18-JUN-1999;	99US-0139454.	
PR	18-JUN-1999;	99US-0139455.	
PR	18-JUN-1999;	99US-0139456.	
PR	18-JUN-1999;	99US-0139457.	
PR	18-JUN-1999;	99US-0139458.	
PR	18-JUN-1999;	99US-0139459.	
PR	18-JUN-1999;	99US-0139460.	
PR	18-JUN-1999;	99US-0139461.	
PR	18-JUN-1999;	99US-0139462.	
PR	18-JUN-1999;	99US-0139463.	
PR	18-JUN-1999;	99US-0139750.	
PR	18-JUN-1999;	99US-0139763.	
PR	21-JUN-1999;	99US-0139817.	
PR	22-JUN-1999;	99US-0139899.	
PR	23-JUN-1999;	99US-0140353.	
PR	23-JUN-1999;	99US-0140354.	
PR	24-JUN-1999;	99US-0140695.	
PR	28-JUN-1999;	99US-0140823.	
PR	29-JUN-1999;	99US-0140991.	
PR	30-JUN-1999;	99US-0141287.	
PR	01-JUL-1999;	99US-0141842.	
PR	01-JUL-1999;	99US-0142154.	
PR	02-JUL-1999;	99US-0142055.	
PR	06-JUL-1999;	99US-0142390.	
PR	08-JUL-1999;	99US-0142803.	
PR	09-JUL-1999;	99US-0142920.	
PR	12-JUL-1999;	99US-0142977.	
PR	13-JUL-1999;	99US-0143542.	
PR	14-JUL-1999;	99US-0143624.	
PR	15-JUL-1999;	99US-0144005.	
PR	16-JUL-1999;	99US-0144085.	
PR	16-JUL-1999;	99US-0144086.	
PR	19-JUL-1999;	99US-0144325.	
PR	19-JUL-1999;	99US-0144331.	
PR	19-JUL-1999;	99US-0144332.	
PR	19-JUL-1999;	99US-0144333.	
PR	19-JUL-1999;	99US-0144334.	
PR	19-JUL-1999;	99US-0144335.	
PR	20-JUL-1999;	99US-0144352.	
PR	20-JUL-1999;	99US-0144632.	
PR	20-JUL-1999;	99US-0144884.	
PR	21-JUL-1999;	99US-0144814.	
PR	21-JUL-1999;	99US-0145086.	
PR	21-JUL-1999;	99US-0145088.	
PR	22-JUL-1999;	99US-0145085.	
PR	22-JUL-1999;	99US-0145087.	
PR	22-JUL-1999;	99US-0145089.	
PR	22-JUL-1999;	99US-0145192.	
PR	23-JUL-1999;	99US-0145145.	
PR	23-JUL-1999;	99US-0145218.	
PR	23-JUL-1999;	99US-0145224.	
PR	26-JUL-1999;	99US-0145276.	

PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match 22.3%; Score 27.4; DB 21; Length 1091;
 Best Local Similarity 57.6%; Pred. No. 4.7;
 Matches 49; Conservative 0; Mismatches 36; Indels 0; Gaps

OY	23	GTGGTCGAAAGGCCCAACACTGTCTTCCAGTCAGTGTAGTGTACAGAACGCGT 82
DB	701	GAGATTGAACAAACAGAGAGACATGCTCTTTATAAGTTGGAGAGATGGGAT 760
OY	83	AGCACTAGCGCTTGACAGAACTCA 107
DB	761	ATCGCTATGGCTGCACCTCTACCTCA 785

RESULT 13
 AAF59609
 ID AAF59609 standard; cDNA; 1689 BP.
 XX AAF59609;
 AC AAF59609;
 DT 24-APR-2001 (first entry)
 XX Human cell cycle and proliferation protein CCYPR-20 cDNA, SEQ ID NO:74.
 XX Cell cycle and proliferation protein; CCYPR; human; agonist;
 KW antagonist; gene therapy; detection; gene therapy;
 KW transgenic animal disease model; immune disorder;
 KW developmental disorder; cell signalling disorder;
 KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
 KW arteriosclerosis; asthma; allergy; diabetes mellitus;
 KW menstrual cycle disorder; bacterial infection; ss.
 XX Homo sapiens.
 XX WO200107471-A2.
 XX 01-FEB-2001.
 XX 21-JUL-2000; 2000WO-US19948.
 XX 21-JUL-1999; 99US-0145075.
 XX 08-SEP-1999; 99US-0153129.
 XX 10-NOV-1999; 99US-0164647.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;
 PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
 WPI; 2001-112727/12.
 XX P-PSDB; AAB60472.
 XX Human cell cycle and proliferation proteins and polynucleotides are
 PT used to treat, diagnose and prevent immune, developmental and cell
 PT signaling disorders and cell proliferative disorders including cancer
 XX Claim 5; Page 180; 205pp; English.
 XX Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human
 CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.
 CC CCYPR and agonists of CCYPR are used to treat diseases or conditions
 CC associated with decreased expression of functional CCYPR, while CCYPR
 CC antagonists are used to treat diseases or conditions associated with
 CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies

CC represent PCR primers for a mouse ischaemic condition related sequence,
 CC which are used in the exemplification of the present invention.

XX
 SQ Sequence 363 BP; 115 A; 57 C; 93 G; 98 T; 0 other;

Query Match 22.0%; Score 27; DB 24; Length 363;

Best Local Similarity 66.1%; Pred. No. 4.4;

Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 8 AAAAGTTTGTCTAGTGGTCGAAGGCCCAACACTGTGTCTTCCACGTGAGTTAGGT 66

||| ||||| ||||| || || || ||||| ||||| ||||| ||||| ||

Db 187 AAAGGTTTGTGCAAGTGATGATGCAGAAAAATTGTCTGCTGGGTCTCTGT 245

Search completed: October 26, 2002, 04:00:35
 Job time : 183 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 26, 2002, 03:01:19 ; Search time 1846 Seconds
(without alignments)
1394.347 Million cell updates/sec

Title: US-09-717-321a-1

Perfect score: 123

Sequence: 1 caattgaaaaagtgttgc.....ctcagagaccacaaagtacc 123

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

1	123	100.0	123	6	AX163737	AX163737 Sequence	
c	2	123	100.0	348	6	AX163753	AX163753 Sequence
3	123	100.0	1266	6	AX163738	AX163738 Sequence	
c	4	107.6	87.5	2319	10	BC003828	BC003828 Mus muscu
c	5	107.6	87.5	269081	2	AC068493	AC068493 Mus muscu
c	6	79.6	64.7	352	6	AX163754	AX163754 Sequence
c	7	79.6	64.7	455	11	G26995	G26995 human SRS S
c	8	79.6	64.7	1232	9	HUMPO2ST9	D25274 Homo sapien
c	9	79.6	64.7	2315	9	AK054993	AK054993 Homo sapi
c	10	79.6	64.7	28567	9	HSAL32695	AK132695 Homo sapi
c	11	79.6	64.7	212827	9	AC009412	AC009412 Homo sapi
c	12	77.6	63.1	262	11	G31709	G31709 SWS2233 E
c	13	74.8	60.8	137625	9	AC104663	AC104663 Homo sapi
c	14	74.8	60.8	159960	30	AC024963	AC024963 Homo sapi
c	15	48	39.0	101584	9	CNS01DS5	AL121655 BAC seque
c	16	48	39.0	155943	9	AC012364	AC012364 Homo sapi
c	17	38.2	31.1	110816	9	AC002404	AC002404 Human Chr
c	18	38.2	31.1	186218	2	AC079169	AC079169 Homo sapi
c	19	33.2	27.0	1485	5	AF029250	AF029250 Danio rer
c	20	32.4	26.3	110000	2	AC107387_0	AC107387 Homo sapi
c	21	32.4	26.3	152747	9	AC092576	AC092576 Homo sapi
c	22	32.4	26.3	322593	2	AC107084	AC107084 Homo sapi
c	23	31.4	25.5	605	6	AX341407	AX341407 Sequence
c	24	31.2	25.4	90083	2	AC103167	AC103167 Rattus no
c	25	30.8	25.0	97955	2	AC108543	AC108543 Rattus no
c	26	30.8	25.0	166081	2	AL626786	AL626786 Mus muscu
c	27	30.6	24.9	167100	9	AC008875	AC008875 Homo sapi
c	28	30.6	24.9	236029	2	AC019232	AC019232 Homo sapi
c	29	30.4	24.7	99098	2	AC095793	AC095793 Rattus no
c	30	30.4	24.7	147098	2	AC097183	AC097183 Rattus no
c	31	30.2	24.6	253	11	AU026917	AU026917 Rattus no
c	32	30	24.4	1286	8	EGU37794	U37794 Eucalyptus
c	33	30	24.4	86945	2	AC095700	AC095700 Rattus no
c	34	30	24.4	91940	9	AL445674	AL445674 Human DNA
c	35	30	24.4	183378	2	AL607151	AL607151 Mus muscu
c	36	30	24.4	224712	2	AL606969	AL606969 Mus muscu
c	37	29.8	24.2	307	11	HSB016XG9	253093 H.sapiens (
c	38	29.6	24.1	68108	2	AC090474	AC090474 Homo sapi
c	39	29.4	23.9	509	11	G63194	G63194 SHGC-140826
c	40	29.4	23.9	82024	2	AC023210	AC023210 Homo sapi
c	41	29.4	23.9	143802	9	AL161431	AL161431 Human DNA
c	42	29.4	23.9	175504	9	AC019196	AC019196 Homo sapi
c	43	29.4	23.9	273403	9	AC011498	AC011498 Homo sapi
c	44	29.2	23.7	124645	9	HUAC004638	AC004638 Homo sapi
c	45	29.2	23.7	124998	9	AC026397	AC026397 Homo sapi

ALIGNMENTS

RESULT 1	AX163737	AX163737	Sequence 1 from Patent WO0138579.	123_bp	DNA	linear	PAT 22-JUN-2001
LOCUS	AX163737	AX163737	AX163737				
DEFINITION	AX163737	AX163737	AX163737				
ACCESSION	AX163737	AX163737	AX163737				
VERSION	AX163737.1	AX163737.1	GI:14544843				
KEYWORDS							
SOURCE	Norway rat.						
ORGANISM	Rattus norvegicus						
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;						
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;						
	Rattus.						
REFERENCE	1 (bases 1 to 123)						
AUTHORS	Gould-Rothberg, B.E., Dipippo, V.A., Ramsch, T.M. and Gerwein, R.W.						
TITLE	Method of identifying toxic agents using nsaid-induced differential gene expression in liver						
JOURNAL	Patent: WO 0138579-A 1 31-MAY-2001;						
	Curagen Corporation (US)						
FEATURES	Location/Qualifiers						
source	1..123						
	/organism="Rattus norvegicus"						
	/db_xref="taxon:10116"						
BASE COUNT	36 a	28 c	30 g	29 t			

ORIGIN
Query Match 100.0%; Score 123; DB 6; Length 123;
Best Local Similarity 100.0%; Pred. No. 6.9e-33;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAATTGAAAAGTTTGTCTAGTGGTGAAGGCCCAACACTGTGTCTTCCAGTGAG 60
|||||
Db 1 CAATTGAAAAGTTTGTCTAGTGGTGAAGGCCCAACACTGTGTCTTCCAGTGAG 60
|||||

Qy 61 TTAGTTGTACAGAACGGCGTTAGCTAGCGTTGACAGAACCTTCACAGACCCAAAGGT 120
|||||
Db 61 TTAGTTGTACAGAACGGCGTTAGCTAGCGTTGACAGAACCTTCACAGACCCAAAGGT 120
|||||

Qy 121 ACC 123
|||
Db 121 ACC 123

RESULT 2
AX163753/c
LOCUS 163753
DEFINITION Sequence 17 from Patent W00138579.
ACCESSION AX163753
VERSION AX163753.1 GI:14544859
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 348)
AUTHORS Gould-Rothberg, B.E., Dippio, V.A., Ramse, T.M. and Gerwein, R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential
gene expression in liver
JOURNAL Patent: WO 0138579-A 17 31-MAY-2001;
Curagen Corporation (US)
FEATURES
source 1. 348
/db_xref="taxon:10116"
BASE COUNT 103 a 76 c 67 g 102 t
ORIGIN

Query Match 100.0%; Score 123; DB 6; Length 348;
Best Local Similarity 100.0%; Pred. No. 7.6e-33;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAATTGAAAAGTTTGTCTAGTGGTGAAGGCCCAACACTGTGTCTTCCAGTGAG 60
|||||
Db 212 CAATTGAAAAGTTTGTCTAGTGGTGAAGGCCCAACACTGTGTCTTCCAGTGAG 153
|||||

Qy 61 TTAGTTGTACAGAACGGCGTTAGCTAGCGTTGACAGAACCTTCACAGACCCAAAGGT 120
|||||
Db 152 TTAGTTGTACAGAACGGCGTTAGCTAGCGTTGACAGAACCTTCACAGACCCAAAGGT 93
|||||

Qy 121 ACC 123
|||
Db 92 ACC 90

RESULT 3
AX163738
LOCUS 1266 bp
DEFINITION Sequence 2 from Patent W00138579.
ACCESSION AX163738
VERSION AX163738.1 GI:14544844
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 1266)
AUTHORS Gould-Rothberg, B.E., Dippio, V.A., Ramse, T.M. and Gerwein, R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential
gene expression in liver
JOURNAL Patent: WO 0138579-A 2 31-MAY-2001;
Curagen Corporation (US)
FEATURES
source 1. 1266
/db_xref="taxon:10116"
BASE COUNT 385 a 258 c 285 g 338 t
ORIGIN

Query Match 100.0%; Score 123; DB 6; Length 1266;
Best Local Similarity 100.0%; Pred. No. 8.6e-33;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAATTGAAAAGTTTGTCTAGTGGTGAAGGCCCAACACTGTGTCTTCCAGTGAG 60
|||||
Db 146 CAATTGAAAAGTTTGTCTAGTGGTGAAGGCCCAACACTGTGTCTTCCAGTGAG 205
|||||

Qy 61 TTAGTTGTACAGAACGGCGTTAGCTAGCGTTGACAGAACCTTCACAGACCCAAAGGT 120
|||||
Db 206 TTAGTTGTACAGAACGGCGTTAGCTAGCGTTGACAGAACCTTCACAGACCCAAAGGT 265
|||||

Qy 121 ACC 123
|||
Db 266 ACC 268

RESULT 4
BC003828/c
LOCUS 2319 bp
DEFINITION Mus musculus, ras-related C3 botulinum toxin substrate 1 (rho
family, small GTP binding protein rac1), clone MGC:6235
IMAGE:3593957, mRNA, complete cds.
ACCESSION BC003828
VERSION BC003828.1 GI:13277917
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2319)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NCI-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapsb-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) medepaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 11 Row: e Column: 22
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 53885.
FEATURES
source Location/Qualifiers
1. 2319
/db_xref="taxon:10090"

```
/clones="MGC:6235 IMAGE:3593957"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha
model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI_CGAP_Mam1"
/lab_host="DH10B"
/notes="Vector: PCMV-SPORT6"
198. .776
/codon_start=1
/product="ras-related C3 botulinum toxin substrate 1 (rho
family, small GTP binding protein Rac1)"
/protein_id="AA03828.1"
/db_xref="GI:13277918"
/translacion="MQAIKCVVGDGAVGKTCLLISVTTNAPGEYIPTVDNYSANV
HVDKPNVLGLMDPAGDEDYDLRPLSPQDTDFELICFSLVSPASFNRAKWTPEYR
HHCPTNLLVGTGLDLDKDDKDTTEKLRKLLPITVPOGLAMAKEIGAVKYLECSAL
TORGLKRVFDEAIRAVLCPPVKKRKKLL"
BASE COUNT 603 a 573 c 503 g 640 t
ORIGIN
Query Match 87.5%; Score 107.6; DB 10; Length 2319;
Best Local Similarity 92.6%; Pred. No. 2.7e-27;
Matches 113; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 CAATTGAAAAAGTTGTTCTAGTGGTGAAGAGCCCAACACTGTTCTTCCAGTGAG 60
Db 2173 CAATTGAAAAAGTTGTTCTAGTGGTGAAGAGCCCAACACTGTTCTTCCAGTGAG 2114
QY 61 TTAGTGTGTACAGAGCGGTTAGTACACTAGGCTGTGACAGAACCTTCACAGACCCCAAGGT 120
Db 2113 TTAGTGTGTACAGAGCGGTTAGTACACTAGGCTGTGACAGAACCTTCACAGACCCCAAGGA 2054
QY 121 AC 122
Db 2053 AC 2052
RESULT 5
AC068493/c
LOCUS 269081 bp DNA linear HTG 25-JAN-2002
DEFINITION Mus musculus clone RP23-76K1 strain C57BL6/J, WORKING DRAFT
SEQUENCE, 26 unordered pieces.
ACCESSION AC068493 GI:15148081
VERSION AC068493.10
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
High Throughput Mouse Sequencing
Unpublished
JOURNAL 2 (bases 1 to 269081)
REFERENCE Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
Direct Submission
Submitted (03-MAY-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
COMMENT On Aug 11, 2001 this sequence version replaced gi:14993654.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site:
http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/sequence/mous
e.html
Contact: gntm@capecod.bwh.harvard.edu
-----Summary Statistics
Center project name: ABN
Sequencing vector: pUC18; L08752
```

```
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 256302 at least Q20
*Consensus quality: 254124 at least Q30
*Consensus quality: 250276 at least Q40
*Estimated insert size: agarose-FP - N/A
**Estimated insert size: 268581 - sum-of-contigs
Quality coverage: agarose-FP - N/A
Quality/cov: 6.2 x in Q20 bases; sum-of-contigs estimation
-----
NOTE: This is a 'working draft' sequence. It currently
consists of 26 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 51730: contig of 51730 bp in length
* 51731 51750: gap of unknown length
* 51751 82423: contig of 30673 bp in length
* 82424 82443: gap of unknown length
* 82444 118158: contig of 35715 bp in length
* 118159 118178: gap of unknown length
* 118179 138441: contig of 21263 bp in length
* 138442 139461: gap of unknown length
* 139462 160638: contig of 21177 bp in length
* 160639 160658: gap of unknown length
* 160659 177611: contig of 16953 bp in length
* 177612 177632: gap of unknown length
* 177632 190991: contig of 13360 bp in length
* 190992 191011: gap of unknown length
* 191012 200128: contig of 9117 bp in length
* 200129 200148: gap of unknown length
* 200149 211882: contig of 11734 bp in length
* 211883 211902: gap of unknown length
* 211903 220773: contig of 8871 bp in length
* 220774 220793: gap of unknown length
* 220794 229643: contig of 8850 bp in length
* 229644 229663: gap of unknown length
* 229664 238339: contig of 8676 bp in length
* 238340 238359: gap of unknown length
* 238360 24848: contig of 6489 bp in length
* 24849 24868: gap of unknown length
* 24869 249725: contig of 4857 bp in length
* 249726 249745: gap of unknown length
* 249746 252523: contig of 2778 bp in length
* 252524 252543: gap of unknown length
* 252544 257520: contig of 4977 bp in length
* 257521 257540: gap of unknown length
* 257541 260397: contig of 2856 bp in length
* 260397 260416: gap of unknown length
* 260417 262709: contig of 2293 bp in length
* 262710 262729: gap of unknown length
* 262730 263934: contig of 1205 bp in length
* 263935 263954: gap of unknown length
* 263955 264196: contig of 242 bp in length
* 264197 264216: gap of unknown length
* 264217 264629: contig of 413 bp in length
* 264630 264649: gap of unknown length
* 264650 265608: contig of 959 bp in length
* 265609 265628: gap of unknown length
* 265629 266935: contig of 1307 bp in length
* 266936 266955: gap of unknown length
* 266956 268335: contig of 1380 bp in length
* 268336 268355: gap of unknown length
* 268356 268369: contig of 194 bp in length
* 268370 268550: gap of unknown length
* 268551 269081: contig of 512 bp in length.
Location/Qualifiers
1..269081
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
```

/clone="RP23-76K1"
/sex="male"
1. .51730
/note="assembly_name:Contig138"
51751. .82423
/note="assembly_name:Contig137"
82444. .118158
/note="assembly_name:Contig136"
118179. .139441
/note="assembly_name:Contig135"
139462. .160638
/note="assembly_name:Contig134"
160659. .177611
/note="assembly_name:Contig133"
177632. .190991
/note="assembly_name:Contig132"
191012. .200128
/note="assembly_name:Contig131"
200149. .211882
/note="assembly_name:Contig130"
211903. .220773
/note="assembly_name:Contig129"
220794. .229643
/note="assembly_name:Contig128"
229664. .238339
/note="assembly_name:Contig127"
238360. .244848
/note="assembly_name:Contig126"
244869. .249725
/note="assembly_name:Contig125"
249746. .252523
/note="assembly_name:Contig124"
vector_side:left
252544. .257520
/note="assembly_name:Contig123"
257541. .260396
/note="assembly_name:Contig122"
260417. .262709
/note="assembly_name:Contig121"
262730. .263934
/note="assembly_name:Contig120"
263955. .264196
/note="assembly_name:Contig119"
264217. .264629
/note="assembly_name:Contig118"
264650. .265608
/note="assembly_name:Contig117"
265629. .266935
/note="assembly_name:Contig116"
266956. .268335
/note="assembly_name:Contig115"
268356. .268549
/note="assembly_name:Contig114"
268570. .269081
/note="assembly_name:Contig113"
BASE COUNT 69726 a 66027 c 64672 g 67724 t 932 others
ORIGIN

Query Match 87.5%; Score 107.6; DB 2; Length 269081;
Best Local Similarity 92.6%; Pred. No. 4.2e-27;
Matches 113; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CAATTGAAAGAACTTTGTTCTAGCTGTCGAAAGCCCAACACGTGTTCTTCCAGTGAG 60
|||||
Db 63360 CAATTGAAAGAACTTTGTTCTAGCTGTCGAAAGCCCAACACGTGTTCTTCCAGTGAG 63301
|||||

QY 61 TTAGGTTGTACAGACGGCTTAGCAGCTAGCAGTGTACAGACCTTCACAGACCCAAAGGT 120
|||||
Db 63300 TTAGGTTGTACAGACGGCTTAGCAGCTAGCAGTGTACAGACCTTCACAGACCCAAAGGT 63241
|||||

QY 121 AC 122
||

Db 63240 AC 63239

RESULT 6
LOCUS AX163754/c
DEFINITION Sequence 18 from Patent WO0138579.
ACCESSION AX163754
VERSION AX163754.1 GI:14544860
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 352)
AUTHORS Gould-Rothberg,B.E., Dippio,V.A., Ransoh,J.M. and Gerweil,R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential
gene expression in liver
JOURNAL Patent: WO 0138579-A 18 31-MAY-2001;
Curagen Corporation (US)
FEATURES
source Location/Qualifiers
1. .352
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 86 a 77 c 72 g 117 t
ORIGIN

Query Match 64.7%; Score 79.6; DB 6; Length 352;
Best Local Similarity 83.1%; Pred. No. 2e-17;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAGAACTTTGTTCTAGCTGTCGAAAGCCCAACACGTGTTCTTCCAGTGAG 60
|||||
Db 216 CAATTGAAAGAACTTTGTTCTAGCTGTCGAAAGCCCAACACGTGTTCTTCCAGTGAG 157
|||||

QY 61 TTAGGTTGTACAGACGGCTTAGCAGCTAGCGC--TTGACAGAACCTTCACAGACCCAAAG 118
|||||
Db 156 TTAAGTTGTACAGACATCGTCAGCAGCTAGCAGTGTACAGAACCTTCACAGACCCAAAG 97
|||||

QY 119 GTAC 122
||

Db 96 GAAC 93

RESULT 7
LOCUS G26995
DEFINITION human STS SHGC-31947, sequence tagged site.
ACCESSION G26995
VERSION G26995.1 GI:1375245
KEYWORDS STS: STS sequence; primer; sequence tagged site.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 455)
AUTHORS Myers,R.M.
JOURNAL Unpublished
COMMENT

Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu

Primer A: AAGTCATACATGCAAAACATATGCG
Primer B: GCGGAGATACAGCGTGG
STS size: 131
PCR Profile:
Initial incubation: 94 degrees C for 90 seconds

construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.): 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES

source

Location/Qualifiers

1..2315
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="BRACE2008958"
/tissue_type="cerebellum"
/clone_lib="BRACE2"
/notes="cloning vector: pME18SFL3"
BASE COUNT 603 a 531 c 501 g 680 t
ORIGIN

Query Match 64.7%; Score 79.6; DB 9; Length 2315;

Best Local Similarity 83.1%; Pred. No. 2.4e-17;

Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

1 CAATTGAAAGAGTTCTTCTAGTGGTCGAAAGGCCCAACACGTGTCTTCCAGTGAG 60

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 2219 CAATTGAAAGAGTTCTTCTAGTGGTCGAAAGGCCCAACACGTGTCTTCCAGTGAG 2160

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 61 TTAGTTGTACAGACGGCTTACGACTAGCGC--TTGACAGAACCTTCACAGACCCCAAAG 118

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 2159 TTAAGTTGTACAGACATCTGACACTACGACAGTTTACAGAACCTTCACAGACCCCAAAG 2100

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 119 GTAC 122

|||||

DB 2099 GAAC 2096

|||||

RESULT 10 HSA132695/c 28567 bp DNA linear PRI 09-NOV-2000

LOCUS HSA132695 Homo sapiens rac1 gene.

DEFINITION AJ132695

ACCESSION AJ132695

VERSION AJ132695.5 GI:8574037

KEYWORDS alternative splicing; Alu; AluJo; AluSg1; AluSp; AluSx; Alu-rich; Cyt-rich; Mix; rac1 gene; rac1 protein; Rac1b protein; repetitive sequence.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Matos, P., Skaug, J., Marques, B., Beck, S., Verissimo, F., Gespach, C., Boavida, M.G., Scherer, S.W. and Jordan, P.

TITLE Small GTPase Rac1: Structure, Localization, and Expression of the Human Gene

JOURNAL Biochem. Biophys. Res. Commun. 277 (3), 741-751 (2000)

REFERENCE 11062023

AUTHORS Jordan, P.

TITLE Direct Submission

JOURNAL Submitted (03-PEB-1999) Jordan P., Centro de Genetica Humana, Laboratorio de Oncobiologia, Instituto Nacional de Saude 'Dr. Ricardo Jorge', Avenida Padre Cruz, 1649-016 Lisboa, PORTUGAL

COMMENT On Jun 20, 2000 this sequence version replaced gi:7248282. Related sequence AI638561.

FEATURES

source

1..28567

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7p22"

/clone="H. NH0425A05, RPCI-11"

repeat_region 1..109

/rpt_family="AluSg"

repeat_region 125..422

/rpt_family="AluSc"

repeat_region 441..742

/rpt_family="AluSq"
743..1226
/gene="rac1"
743..28117
/gene="rac1"
911..983
/rpt_family="GC_rich"
936
/gene="rac1"
/note="alternative transcription start site"
1026
/gene="rac1"
/note="alternative transcription start site"
1037
/gene="rac1"
/note="alternative transcription start site"
1053
/gene="rac1"
/note="alternative transcription start site"
1106..1175
/rpt_family="GC_rich"
Join(1227..1261,13701..13772,18413..18530,22804..22860,
24268..24330,26010..26169,26458..26588)
/gene="rac1"
/note="alternative"
/codon_start=1
/product="Rac1b protein"
/protein_id="CAA10733.6"
/db_xref="GI:8574039"
/db_xref="SPTREMBL:O95501"
/translation="MQAIKCVVVDGAVGKTCLLISYTTNAPGCEIPTVFDNYSANV
MVDGKPNVLGLWDTAGQEDYDLRLPLSYDPTGKDTSRGKDKPIADVPLICFS
LVSPSFENVRKWTPEVHHCPTPIILVGTGLDLRDDDTIEKLKKLLTPTTYPQGLA
MAKEIGAVKYLECSALTQRLKTVDEAIRAVLCPPPPVKKRKKLL"
<1227..1261
/number=1
Join(1227..1261,13701..13772,18413..18530,24268..24330,
26010..26169,26458..26588)
/gene="rac1"
/codon_start=1
/product="Rac1 protein"
/protein_id="CAB53579.5"
/db_xref="GI:8574038"
/translation="MQAIKCVVVDGAVGKTCLLISYTTNAPGCEIPTVFDNYSANV
MVDGKPNVLGLWDTAGQEDYDLRLPLSYDPTGKDTSRGKDKPIADVPLICFS
LVSPSFENVRKWTPEVHHCPTPIILVGTGLDLRDDDTIEKLKKLLTPTTYPQGLA
MAKEIGAVKYLECSALTQRLKTVDEAIRAVLCPPPPVKKRKKLL"
1262..13700
/gene="rac1"
/number=1
1390..1419
/rpt_family="GC_rich"
1752..1793
/rpt_family="FLAM"
1794..2005
/rpt_family="AluJo"
2140..2171
/rpt_family="Simple_repeat"
2666..2972
/rpt_family="AluSx"
2980..3267
/rpt_family="AluSx"
3268..3289
/rpt_family="(T)n"
3802..4102
/rpt_family="AluJb"
4122..4386
/rpt_family="AluSx"
4538..4719
/rpt_family="LTR33A"
4754..5053
/rpt_family="AluSq"

```

repeat_region 5297..5594
/rpt_family="AluSx"
repeat_region 5655..5757
/rpt_family="AluSx"
repeat_region 5758..6062
/rpt_family="AluSx"
repeat_region 6063..6198
/rpt_family="AluSx"
repeat_region 6222..6401
/rpt_family="AluSg/x"
repeat_region 7162..7238
/rpt_family="L1MC5"
repeat_region 7348..7367
/rpt_family="Simple_repeat"
repeat_region 7368..7656
/rpt_family="AluSx"
repeat_region 7880..8183
/rpt_family="AluSx"
repeat_region 8484..8778
/rpt_family="AluSx"
repeat_region 9071..9365
/rpt_family="AluSx"
repeat_region 10279..10583
/rpt_family="AluSp"
repeat_region 10609..10864
/rpt_family="L2"
repeat_region 10970..11276
/rpt_family="AluJo"
repeat_region 11276..11594
/rpt_family="AluJb"
repeat_region 11744..11825
/rpt_family="L2"
repeat_region 11764..11987
/rpt_family="MIR"
repeat_region 12989..13307
/rpt_family="AluSx"
exon 13701..13772
/gene="rac1"
Intron 13773..18412
/number=2
/gene="rac1"
repeat_region 13954..14235
/rpt_family="AluSx"
repeat_region 14807..14831
/rpt_family="Simple_repeat"
repeat_region 14836..15128
/rpt_family="AluSx"
repeat_region 15408..15715
/rpt_family="AluSx"
repeat_region 15872..15896
/rpt_family="AT-rich"
repeat_region 15902..16186
/rpt_family="AluSx"
repeat_region 16187..16314
/rpt_family="AluJo"
repeat_region 16413..16585
/rpt_family="Simple_repeat"
repeat_region 16586..16679
/rpt_family="FLAM_A"
repeat_region 16707..16810
/rpt_family="L1M4"
repeat_region 16811..17108
/rpt_family="AluSg"
repeat_region 17109..17345
/rpt_family="L1M4"
repeat_region 17377..17686
/rpt_family="AluSx"
repeat_region 17703..17729
/rpt_family="Simple_repeat"
exon 18413..18530
/gene="rac1"
/number=3

Query Match 64.7% Score 79.6; DB 9; Length 28567;
Best Local Similarity 83.1%; Pred. No. 3e-17;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAAAGTTTCTAGTGGTCGAAAGCCCAACACTGTGTCTTGGCAGTGAG 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 28012 CAATTAAAAAATTCGTCTAGTGGCTGAAGGTCCTCCACGCTGTATTCTGCCAGTGAG 27953

QY 61 TTAGGTTGTACAGAACGGCGTTAGCTAGTGGC--TTGACAGAACCTCAGACCCCAAG 118
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27952 TTAAGTTGTACAGAACATCGTCAGCTAGCAGACAGTTTACAGAACCTCAGACCCCAAG 27893

QY 119 GTAC 122
Db 27892 GAAC 27889

RESULT 11
AC009412/c 212827 bp DNA linear PRI 07-NOV-2001
LOCUS AC009412 Homo sapiens BAC clone RP11-425P5 from 7, complete sequence.
DEFINITION AC009412
ACCESSION AC009412
VERSION AC009412.6 GI:14190769
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 212827)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 212827)
AUTHORS Hou,S., Maupin,R., Haakenson,W., Gregory,S. and Belter,E.
TITLE The sequence of Homo sapiens BAC clone RP11-425P5
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 212827)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 212827)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 212827)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 23, 2001 this sequence version replaced gi:13431187.
----- Genome Center

```

Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@wustl.edu
 ----- Summary Statistics -----
 Center project name: H_NH0425P05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
 VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is CTD-2195F2, 200 bp overlapp. Actual start of this clone is at base position 1 of RP11-425P5; actual end is at base position 213633 of RP11-425P5.

FEATURES

source

1. .212827
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7"

/clone="RP11-425P5"

/clone_lib="RP11-11"

1623..1696

/rpt_family="MIR"

1698..2006

/rpt_family="Alu"

2024..2337

/rpt_family="Alu"

3006..3093

/rpt_family="L1"

3453..3552

/rpt_family="L2"

4924..5046

/rpt_family="ERV1"

5318..5693

/note="similar to EST BF996284 (NID:g12402607)"

5580..5845

/rpt_family="Alu"

5925..6277

/rpt_family="MIR"

6842..7041

/note="similar to EST BF369507 (NID:g11331532)"

repeat_region	7364..7661	/rpt_family="Alu"	
repeat_region	8029..8048	/rpt_family="(CAAAA)n"	
repeat_region	8379..8678	/rpt_family="Alu"	
misc_feature	8560..8762	/note="similar to EST AA699398 (NID:g2702592) z140a03.s1"	
repeat_region	8720..9031	/rpt_family="Alu"	
misc_feature	9083..9560	/note="similar to EST AA633603 (NID:g2556817) ae66b02.s1"	
repeat_region	9598..9899	/rpt_family="Alu"	
misc_feature	9871..10098	/note="similar to EST AA699398 (NID:g2702592) z140a03.s1"	
repeat_region	10746..11044	/rpt_family="L1"	
repeat_region	11075..11101	/rpt_family="AT-rich"	
repeat_region	11102..11372	/rpt_family="Alu"	
repeat_region	11411..11465	/rpt_family="GA-rich"	
repeat_region	11552..11644	/rpt_family="L2"	
repeat_region	11685..12193	/rpt_family="L1"	
repeat_region	12194..12486	/rpt_family="Alu"	
repeat_region	12487..13003	/rpt_family="L1"	
repeat_region	13004..13305	/rpt_family="Alu"	
repeat_region	13306..13501	/rpt_family="L1"	
repeat_region	13503..13892	/rpt_family="L1"	
misc_feature	13893..14224	/note="similar to EST BE061102 (NID:g8405752)"	
repeat_region	14244..14544	/rpt_family="Alu"	
repeat_region	14546..14828	/rpt_family="Alu"	
repeat_region	15091..15189	/rpt_family="MIR"	
repeat_region	15672..16449	/rpt_family="Achobo"	
repeat_region	16640..16844	/rpt_family="Alu"	
repeat_region	16917..16959	/rpt_family="Achobo"	
repeat_region	16960..17287	/rpt_family="Alu"	
repeat_region	17288..17572	/rpt_family="Achobo"	
misc_feature	17610..17938	/note="similar to EST BF996427 (NID:g12402750)"	
repeat_region	17793..18068	/rpt_family="Alu"	
repeat_region	18251..18537	/rpt_family="Alu"	
misc_feature	18485..18884	/note="similar to EST BE062347 (NID:g8406997)"	
misc_feature	18661..18809	/note="similar to EST BG186544 (NID:g13708231)"	
repeat_region	18957..19160	/rpt_family="L1"	
repeat_region	19314..19528	/rpt_family="L1"	
repeat_region	19819..20135	/rpt_family="Alu"	
repeat_region	21101..21269		

```
repeat_region /rpt_family="L1"
21882..22203 /rpt_family="L1"
misc_feature 22429..22797 /note="similar to EST BF993478 (NID:gi23399801)"
repeat_region 22717..23007 /rpt_family="L1"
22993..23191 /note="similar to EST BF767044 (NID:gi21115035)"
repeat_region 23167..23471 /rpt_family="Alu"
23511..23930 /note="similar to EST AW105611 (NID:g6076346) xda9g02.xl"
repeat_region 25387..25677 /rpt_family="Alu"
25706..26013 /rpt_family="Alu"
```

```
Query Match 64.7%; Score 79.6; DB 9; Length 212827;
Best Local Similarity 83.1%; Pred. No. 3.6e-17;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAAAGTTTCTTCTAGTGGTGGAAAGGCCCAACACTGTGTCTTGTCCAGTGAG 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 176040 CAATTTAAAAAATCTGTGTAGTGGCTGAAGGTCCACGCTGTATCTCGCCAGTGAG 175981

QY 61 TTAGGTTGTACAGAACGGCTGTAGCACTAGCGC--TTGACAGAACCTCAGACGCCAAAG 118
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 175980 TTAAGTTGTACAGAACATCTGTAGCACTAGCACTTACAGAACCTCAGACGCCAAAG 175921

QY 119 GTAC 122
| ||
Db 175920 GAAC 175917
```

```
RESULT 12
G31709
LOCUS G31709 262 bp mRNA linear STS 28-SEP-1998
DEFINITION SW522333 Eric D. Green Homo sapiens STS cDNA, sequence tagged site.
ACCESSION G31709
VERSION G31709.1 GI:3660607
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 262)
AUTHORS Bouffard,G.G., Iyer,L.M., Idol,J.R., Braden,V.V., Cunningham,A.F.,
Weintraub,L.A., Mohr-Tidwell,R.M., Peluso,D.C., Fulton,R.S.,
Leckie,M.P. and Green,E.D.
TITLE A collection of 1814 human chromosome 7-specific STSs
JOURNAL Genome Res. 7 (1), 59-64 (1997)
MEDLINE 97189344
PUBMED 9037602
REFERENCE 2 (bases 1 to 262)
AUTHORS Green,E.D.
TITLE Human chromosome 7 STSs (1997)
JOURNAL Unpublished
COMMENT On Sep 29, 1998 this sequence version replaced gi:1916434.
Synonyms: AB4082
Contact: Eric D. Green
Genome Technology Branch
National Human Genome Research Institute/NIH
49 Convent Dr., MSC4431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892
Tel: 3014020201
Fax: 3014024735
Email: egreen@nhgri.nih.gov
Primer A: CACGCTGATTCCTGCCAGT
Primer B: GGACACACGCCCTCTGTAGT
STS size: 123
PCR Profile:
Presoak: 0 degrees C for 0.00 minute(s)
Denaturation: 92 degrees C for 0.17 minute(s)
```

```
Annealing: 55 degrees C for 1.00 minute(s)
Polymerization: 72 degrees C for 1.00 minute(s)
PCR Cycles: 35
Thermal Cycler: PerkinElmer 9600

Protocol:
Template: 30-100 ng
Primer: each 1 uM
dNTPs: each 200 uM
Taq Polymerase: 0.05 units/ul
Total Vol: 10 ul

Buffer:
MgCl2: 1.5 mM
KCl: 100 mM
Tris-HCl: 10 mM
NH4Cl: 5 mM
pH: 8.6
```

This STS has been incorporated into the NHGRI chromosome 7 physical map, but was developed by another investigator. See GenBank record: F03612 For additional information about the NHGRI chromosome 7 mapping project, see <http://www.nhgri.nih.gov/DIR/STB/CHR7>. Also see Genomics 11:548-64 (1991) [MUID=92128937].

```
FEATURES
source          1..262
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"
                 /map="7"
                 /clone_lib="Eric D. Green"
                 134..256
primer_bind     134..153
primer_bind     complement(237..256)
BASE COUNT      88 a 55 c 53 g 64 t 2 others
ORIGIN
```

```
Query Match 63.1%; Score 77.6; DB 11; Length 262;
Best Local Similarity 81.5%; Pred. No. 1e-16;
Matches 101; Conservative 0; Mismatches 21; Indels 2; Gaps 1;

QY 1 CAATTGAAAAAGTTTCTTCTAGTGGTGGAAAGGCCCAACACTGTGTCTTGTCCAGTGAG 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 97 CAATTTAAAAAATCTGTGTAGTGGCTGAAGGTCCACGCTGTATCTCGCCAGTGAG 156

QY 61 TTAGGTTGTACAGAACGGCTGTAGCACTAGCGC--TTGACAGAACCTCAGACGCCAAAG 118
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 157 TTAAGTTGTACAGAACATCTGTAGCACTAGCACTTACAGAACCTCAGACGCCAAAG 216

QY 119 GTAC 122
| ||
Db 217 GAAC 220
```

```
RESULT 13
AC104663/c
LOCUS AC104663 137625 bp DNA linear PRI 10-FEB-2002
DEFINITION Homo sapiens chromosome 4 cDNA clone RP11-713M6, complete sequence.
ACCESSION AC104663 AC024963
VERSION AC104663.3 GI:18643727
KEYWORDS HYG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 137625)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 137625)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
```


Search completed: October 26, 2002, 04:33:19
Job time : 1953 secs